Scaling limits of Markov branching trees

with applications to Galton-Watson and random unordered trees *

Bénédicte Haas[†] and Grégory Miermont[‡]

Abstract

We consider a family of random trees satisfying a Markov branching property. Roughly, this property says that the subtrees above some given height are independent with a law that depends only on their total size, the latter being either the number of leaves or vertices. Such families are parameterized by sequences of distributions on partitions of the integers, that determine how the size of a tree is distributed in its different subtrees. Under some natural assumption on these distributions, stipulating that "macroscopic" splitting events are rare, we show that Markov branching trees admit the so-called self-similar fragmentation trees as scaling limits in the Gromov-Hausdorff-Prokhorov topology.

The main application of these results is that the scaling limit of random uniform unordered trees is the Brownian continuum random tree. This extends a result by Marckert-Miermont and fully proves a conjecture by Aldous. We also recover, and occasionally extend, results on scaling limits of consistent Markov branching model, and known convergence results of Galton-Watson trees towards the Brownian and stable continuum random trees.

1 Introduction and main results

The goal of this paper is to discuss the scaling limits of a model of random trees satisfying a simple Markovian branching property, that was considered in [26], and considered in different forms in a number of places [15, 4, 12]. Markov branching trees are natural models of random trees, defined in terms of discrete fragmentation processes. The laws of these trees are indexed by an integer n giving the "size" of the tree, which leads us to consider two distinct (but related) models, in which the sizes are respectively the number of leaves and the number of vertices. We first provide a slightly informal description of our results.

Let $q = (q_n, n \geq 1)$ be a family of probability distributions, respectively on the set $\mathcal{P}_n$ of partitions of the integer n, i.e. of non-increasing integer sequences with sum n. We assume that q_n does not assign mass 1 to the trivial partition (n):

$$q_n((n)) < 1\,, \qquad \text{for every } n \geq 1\,.$$

In order that this makes sense for $n = 1$, we add an extra "empty partition" $\emptyset$ to $\mathcal{P}_1$.

One constructs a random rooted tree with n leaves according to the following procedure. Start from a collection of n indistinguishable balls, and with probability $q_n(\lambda_1, \ldots, \lambda_p)$, split the collection into p sub-collections with $\lambda_1, \ldots, \lambda_p$ balls. Note that there is a chance $q_n((n)) < 1$ that the collection remains unchanged during this step of the procedure. Then, re-iterate the splitting operation independently for each sub-collection using this time the probability distributions $q_{\lambda_1}, \ldots, q_{\lambda_p}$. If a sub-collection consists of a single ball, it can remain single with probability $q_1((1))$ or get wiped out with probability $q_1(\emptyset)$. We continue the procedure until all the balls are wiped out. There is

*This work is partially supported by ANR-08-BLAN-0190 and ANR-08-BLAN-0220-01
†Ceremade, Université Paris-Dauphine
‡Département de Mathématiques d'Orsay, Université Paris-Sud

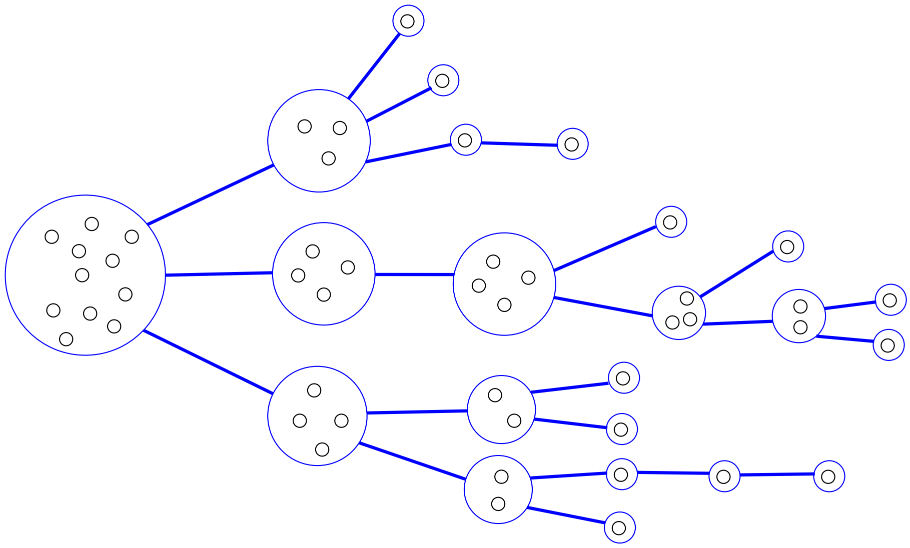

Figure 1: A sample tree T_{11}. The first splitting arises with probability $q_{11}(4,4,3)$.

a natural genealogy associated with this process, which is a tree with n leaves consisting in the n isolated balls just before they are wiped out, and rooted at the initial collection of n balls. See Figure 1 for an illustration. We let P_n^q be the law of this tree.

This construction can be seen as the most general form of *splitting trees* of Broutin & al. [12], and was referred to as trees bearing the so-called *Markov branching property* in [26]. There is also a variant of this procedure that constructs a random tree with n vertices rather than n leaves. This one does not need the hypothesis $q_n((n)) < 1$ for $n \geq 1$, and in fact we only assume $q_1((1)) = 1$ for consistency of the description to follow. Informally, starting from a collection of n balls, we first remove a ball, split the $n-1$ remaining balls in sub-collections with $\lambda_1, \ldots, \lambda_p$ balls with probability $q_{n-1}((\lambda_1, \ldots, \lambda_p))$, and iterate independently on sub-collections until no ball remains. We let Q_n^q be the law of the random tree associated to this procedure.

While most papers so far have been focusing on families of trees having more structure, such as a consistency property when n varies [4, 26, 15] (with the notable exception of Broutin & al. [12]), the main goal of the present work is to study the structure of trees with laws P_n^q or Q_n^q as $n \to \infty$ in a very general situation. The main assumption that we make is that, as $n \to \infty$,

> macroscopic splitting events of the form $n \to (ns_1, ns_2, \ldots) \in \mathcal{P}_n$ for a non-increasing sequence $\mathbf{s} = (s_1, s_2, \ldots)$ with sum 1 and such that $s_1 < 1 - \varepsilon$, for some $\varepsilon \in (0,1)$, are rare events, occurring with probability of order $n^{-\gamma}\nu_\varepsilon(\mathrm{d}\mathbf{s})$ for some $\gamma > 0$, for some finite "intensity" measure ν_ε.

Note that the measures ν_ε should satisfy a consistency property as ε varies, and as ε goes to 0, ν_ε should increase to a possibly infinite measure on the set of non-increasing sequences with sum 1. This means that splitting events that only remove tiny parts from a large collection of balls are allowed to remain more frequent than the order $n^{-\gamma}$. Under this assumption, formalized in hypothesis **(H)** below, we show in Theorem 1 that a tree T_n with law P_n^q, considered as a metric space by viewing its edges as being real segments of lengths of order $n^{-\gamma}$, converges in distribution towards a limiting structure $\mathcal{T}_{\gamma,\nu}$, the so-called self-similar fragmentation tree of [24]:

$$\frac{1}{n^\gamma} T_n \longrightarrow \mathcal{T}_{\gamma,\nu} .$$

When $\gamma \in (0,1)$, a similar result (Theorem 2) holds when T_n has distribution Q_n^q.

The limiting tree $\mathcal{T}_{\gamma,\nu}$ can be seen as the genealogical tree of a continuous model for mass splitting, in some sense analogous to the Markov branching property described above. The above convergence holds in distribution in a space of measured metric spaces, endowed with the so-called Gromov-Hausdorff-Prokhorov topology. This result contrasts with the situation of [12], where it is assumed that macroscopic splitting events occur at every step of the construction. In that case, the height of T_n is of order $\log n$, and no interesting scaling limit exists. A key step in our study will be to use the results from [25], where scaling limits of non-increasing Markov chains were considered: such Markov chains are indeed obtained by considering the successive sizes of collections containing a particular marked ball when going up in the tree T_n.

This general statement allows to recover, and sometimes improve, many results of [26, 27, 33, 14] dealing specifically with Markov branching trees. It also applies to models of random trees that are not *a priori* directly connected to our study. In particular, we recover the results of Aldous [3] and Duquesne [17] showing that the so-called Brownian and stable trees [1, 28, 18, 19] are universal limits for conditioned Galton-Watson trees.

More notably, our results allow to prove that uniform unordered trees with n vertices and degrees bounded by some integer $m+1 \in [3,\infty]$ admit the Brownian continuum random tree as a scaling limit. This was conjectured by Aldous [2] and proved in [29] in the particular case $m = 2$ of a binary branching, using completely different methods from ours. The difficulty of handling such families of random trees comes from the fact that they have no 'nice' probabilistic representations, using for instance branching processes or growth models. As a matter of fact, uniform random unordered trees do not even have the Markov branching property, but turn out to 'almost' bear this property, in a sense that will be explained below.

The rest of this section is devoted to a detailed formalization of our results. Throughout the paper, we use the notations

$$\mathbb{N} = \{1, 2, 3, \ldots\}, \qquad \mathbb{Z}_+ = \{0\} \cup \mathbb{N}, \qquad [n] = \{1, 2, \ldots, n\}, \quad n \in \mathbb{N}.$$

The random variables appearing in this paper are either canonical or defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Acknowledgment. We are deeply indebted to Jean-François Marckert, who suggested that the methods of [26] could be relevant to tackle the problem of scaling limits of uniform unordered trees. This provided the initial spark for the present work and [25].

1.1 Discrete trees

We briefly introduce some formalism for trees. Set $\mathbb{N}^0 = \{\varnothing\}$, and let

$$\mathcal{U} = \bigcup_{n \geq 0} \mathbb{N}^n .$$

For $u = (u_1, \ldots, u_n) \in \mathcal{U}$, we denote by $|u|$ the length of u, also called the height of u. If $u = (u_1, \ldots, u_n)$ with $n \geq 1$, we let $\mathrm{pr}(u) = (u_1, \ldots, u_{n-1})$, and for $i \geq 1$, we let $ui = (u_1, \ldots, u_n, i)$. More generally, for $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_m)$ in $\mathcal{U}$, we let $uv = (u_1, \ldots, u_n, v_1, \ldots, v_m)$ be their concatenation, and for $A \subset \mathcal{U}$ and $u \in \mathcal{U}$, we let $uA = \{uv : v \in A\}$, and simply let $iA = (i)A$ for $i \in \mathbb{N}$. We say that u is a prefix of v if $v \in u\mathcal{U}$, and write $u \preceq v$, defining a partial order on $\mathcal{U}$.

A plane tree is a non-empty, finite subset $\mathbb{t} \subset \mathcal{U}$ (whose elements are called *vertices*), such that

- If $u \in \mathbb{t}$ with $|u| \geq 1$, then $\mathrm{pr}(u) \in \mathbb{t}$,
- If $u \in \mathbb{t}$, then there exists a number $c_u(\mathbb{t}) \in \mathbb{Z}_+$ (the number of children of u) such that $ui \in \mathbb{t}$ if and only if $1 \leq i \leq c_u(\mathbb{t})$.

Let $\partial\mathbb{t} = \{u \in \mathbb{t} : c_u(\mathbb{t}) = 0\}$ be the set of *leaves* of $\mathbb{t}$. If $\mathbb{t}^{(1)}, \dots, \mathbb{t}^{(k)}$ are plane trees, we can define a new plane tree by

$$\langle \mathbb{t}^{(1)}, \dots, \mathbb{t}^{(k)} \rangle = \{\varnothing\} \cup \bigcup_{i=1}^{k} i\mathbb{t}^{(i)} .$$

A plane tree has a natural graphical representation, in which every $u \in \mathbb{t}$ is a vertex, joined to its $c_u(\mathbb{t})$ children by as many edges. But $\mathbb{t}$ carries more information than the graph, as it has a natural ordered structure. In this work, we will not be interested in this order, and we present one way to get rid of this unwanted structure. Let $\mathbb{t}$ be a plane tree, and $\boldsymbol{\sigma} = (\sigma_u, u \in \mathbb{t})$ be a sequence of permutations, respectively $\sigma_u \in \mathfrak{S}_{c_u(\mathbb{t})}$. For $u = (u_1, \dots, u_n) \in \mathbb{t}$, let

$$\boldsymbol{\sigma}(u) = \left(\sigma_\varnothing(u_1), \sigma_{(u_1)}(u_2), \sigma_{(u_1,u_2)}(u_3), \dots, \sigma_{(u_1,\dots,u_{n-1})}(u_n)\right) ,$$

and $\boldsymbol{\sigma}(\varnothing) = \varnothing$. Then the set $\boldsymbol{\sigma}\mathbb{t} = \{\boldsymbol{\sigma}(u) : u \in \mathbb{t}\}$ is a plane tree, obtained intuitively by shuffling the set of children of u in $\mathbb{t}$ according to the permutation σ_u. We say that $\mathbb{t}, \mathbb{t}'$ are equivalent if there exists some $\boldsymbol{\sigma}$ such that $\boldsymbol{\sigma}(\mathbb{t}) = \mathbb{t}'$. Equivalence classes of plane trees will be called (rooted) *unordered trees*, or simply *trees* as opposed to plane trees, and denoted by lowercase letters t.

When dealing with a tree t, we will freely adapt some notations from plane trees when dealing with quantities that do not depend on particular plane representatives. For instance, $\#\mathsf{t}, \#\partial\mathsf{t}$ will denote the number of vertices and leaves of t, while $\varnothing, c_\varnothing(\mathsf{t})$ will denote the root of t and its degree.

We let T be the set of trees, and for $n \geq 1$,

$$\mathsf{T}_n^\partial = \{\mathsf{t} \in \mathsf{T} : \#\partial\mathsf{t} = n\} , \qquad \mathsf{T}_n = \{\mathsf{t} \in \mathsf{T} : \#\mathsf{t} = n\}$$

be the set of trees with n leaves, resp. n vertices. The class of $\{\varnothing\}$ is the vertex tree $\bullet \in \mathsf{T}_1 = \mathsf{T}_1^\partial$.

Heuristically, the information carried in a tree is its graph structure, with a distinguished "root" vertex corresponding to $\varnothing$, and considered up to root-preserving graph isomorphisms — it is not embedded in any space, and its vertices are unlabeled.

It is a simple exercise to see that if $\mathsf{t}^{(i)}, 1 \leq i \leq k$ are trees, and $\mathbb{t}^{(i)}$ is a choice of a plane representative of $\mathsf{t}^{(i)}$ for each i, then the class of $\langle \mathbb{t}^{(i)}, 1 \leq i \leq k \rangle$ does not depend on the particular choice for $\mathbb{t}^{(i)}$. We denote this common class by $\langle \mathsf{t}^{(i)}, 1 \leq i \leq k \rangle$. Note that $j(\mathsf{t}) := \langle \mathsf{t} \rangle$ can be seen as the tree t whose root has been attached to a new root by an edge, and $j^l(\mathsf{t})$, for $l \geq 0$, is tree t whose root has been attached to a new root by a string of l edges. For instance, $j^l(\bullet)$ is the line-tree consisting of a string with length l, rooted at one of its ends. Finally, for trees $\mathsf{t}^{(1)}, \dots, \mathsf{t}^{(k)}$ and $l \geq 1$ we let

$$\langle \mathsf{t}^{(1)}, \dots, \mathsf{t}^{(k)} \rangle_l = j^l(\langle \mathsf{t}^{(1)}, \dots, \mathsf{t}^{(k)} \rangle) ,$$

so $j^l(\bullet) = \langle \bullet \rangle_l$ with these notations.

1.2 Markov branching trees

A partition of an integer $n \geq 1$ is a sequence of integers $\lambda = (\lambda_1, \dots, \lambda_p)$ with $\lambda_1 \geq \dots \geq \lambda_p \geq 1$ and $\lambda_1 + \dots + \lambda_p = n$. The number $p = p(\lambda)$ is called the number of parts of the partition λ, and the partition is called non-trivial if $p(\lambda) \geq 2$. We let $\mathcal{P}_n$ be the set of partitions of the integer n. We also add an extra element $\emptyset$ to $\mathcal{P}_1$, so that $\mathcal{P}_1 = \{(1), \emptyset\}$.

For $\lambda \in \mathcal{P}_n$ and $1 \leq j \leq n$, we define

$$m_j(\lambda) = \#\{i \in \{1, 2, \dots, p(\lambda)\} : \lambda_i = j\} ,$$

the multiplicity of parts of λ equal to j.

By convention, it is sometimes convenient to set $\lambda_i = 0$ for $i > p(\lambda)$, and to identify the sequence λ with the infinite sequence $(\lambda_i, i \geq 1)$. Such identifications will be implicit when needed.

1.2.1 Markov branching trees with a prescribed number of leaves

In this paragraph, the size of a tree $\mathsf{t} \in \mathsf{T}$ is going to be the number $\#\partial\mathsf{t}$ of its leaves.

Let $q = (q_n, n \geq 1)$ be a sequence of probability distributions respectively on $\mathcal{P}_n$,

$$q_n = (q_n(\lambda), \lambda \in \mathcal{P}_n)\,, \qquad \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) = 1\,,$$

such that

$$q_n((n)) < 1\,, \qquad n \geq 1\,. \tag{1}$$

Consider a family of probability distributions $\mathsf{P}^q_n, n \geq 1$ on T^∂_n respectively, such that

1. P^q_1 is the law of the line-tree $\langle \bullet \rangle_G$, where G has a geometric distribution given by

$$\mathbb{P}(G = k) = q_1(\emptyset)(1 - q_1(\emptyset))^k\,, \qquad k \geq 0\,,$$

2. for $n \geq 2$, P^q_n is the law of

$$\langle T^{(i)}, 1 \leq i \leq p(\Lambda) \rangle\,,$$

   where Λ has distribution q_n, and conditionally on the latter, the trees $T^{(i)}, 1 \leq i \leq p(\Lambda)$ are independent with distributions $\mathsf{P}^q_{\Lambda_i}$ respectively.

Alternatively, for $n \geq 2$, P^q_n is the law of $\langle T^{(i)}, 1 \leq i \leq p(\Lambda) \rangle_G$ where G is independent of Λ and geometric with

$$\mathbb{P}(G = k) = (1 - q_n((n)))q_n((n))^k\,, \qquad k \geq 1\,,$$

and conditionally on Λ with law $q_n(\cdot \,|\, \mathcal{P}_n \setminus \{(n)\})$, the trees $T^{(1)}, \ldots, T^{(p(\Lambda))}$ are independent with distributions P_{Λ_i} respectively. A simple induction argument shows that there exists a unique family P^q_n satisfying the properties 1 and 2 above.

A family of random trees $T_n, n \geq 1$ with respective distributions $\mathsf{P}^q_n, n \geq 1$ is called a Markov branching family. The law of the tree T_n introduced in the beginning of the Introduction to describe the genealogy of splitting collections of n balls is P^q_n.

1.2.2 Markov branching trees with a prescribed number of vertices

We now consider the following variant of the above construction, in which the size of a tree t is the number of its vertices. For every $n \geq 1$, let again q_n be a probability distribution on $\mathcal{P}_n$. We do not assume (1), rather, we make the sole assumption that $q_1((1)) = 1$. For every $n \geq 1$, we construct inductively a family of random trees T_n respectively in the set T_n of trees with n vertices, by assuming that for $\lambda = (\lambda_1, \ldots, \lambda_p) \in \mathcal{P}_{n-1}$, with probability $q_{n-1}(\lambda)$, the $n-1$ vertices distinct from the root vertex are dispatched in p subtrees with $\lambda_1 \geq \ldots \geq \lambda_p$ vertices, and that, given these sizes, the p subtrees are independent with same distribution as $T_{\lambda_1}, \ldots, T_{\lambda_p}$ respectively.

Formally,

1. let Q^q_1 be the law of $\bullet$, and
2. for $n \geq 1$, let Q^q_{n+1} be the law of

$$\langle T^{(i)}, 1 \leq i \leq p(\Lambda) \rangle\,,$$

   where Λ has distribution q_n, and conditionally on the latter, the trees $T^{(i)}, 1 \leq i \leq p(\Lambda)$ are independent with distributions $\mathsf{Q}^q_{\Lambda_i}$ respectively.

By induction, these two properties determine the laws $\mathsf{Q}^q_n, n \geq 1$ uniquely.

The construction is very similar to the previous one, and can in fact be seen as a special case, after a simple transformation on the tree; see Section 4.5 below.

1.3 Topologies on metric spaces

The main goal of the present work is to study scaling limits of trees with distributions $\mathsf{P}_n^q, \mathsf{Q}_n^q$, as n becomes large. For this purpose, we need to consider a topological "space of trees" in which such limits can be taken, and define the limiting objects.

A *rooted*[1] metric space is a triple (X, d, ρ), where (X, d) is a metric space and $\rho \in X$ is a distinguished point, called the root. We say that two rooted spaces $(X, \rho, d), (X', \rho', d')$ are isometry-equivalent if there exists a bijective isometry from X onto X that sends ρ to ρ'.

A measured, rooted metric space is a 4-tuple (X, d, ρ, μ), where (X, d, ρ) is a rooted metric space and μ is a Borel probability measure on X. Two measured, rooted spaces (X, d, ρ, μ) and (X, d', ρ', μ') are isometry-equivalent if there exists a root-preserving, bijective isometry ϕ from (X, d, ρ) to (X, d', ρ') such that the push-forward of μ by ϕ is μ'. In the sequel we will almost always identify two isometry-equivalent (rooted, measured) spaces, and will often use the shorthand notation X for the isometry class of a rooted space or a measured, rooted space, in a way that should be clear from the context. Also, if X is such a space and $a > 0$, then we denote by aX the space in which the distance function is multiplied by a.

We denote by $\mathscr{M}$ the set of equivalence classes of compact rooted spaces, and by $\mathscr{M}_{\mathrm{w}}$ the set of equivalence classes of compact measured spaces.

It is well-known (this is an easy extension of the results of [21]) that $\mathscr{M}$ is a Polish space when endowed with the so-called pointed Gromov-Hausdorff distance d_{GH}, where by definition the distance $d_{\mathrm{GH}}((X, \rho), (X', \rho'))$ is equal to the infimum of the quantities

$$\mathrm{dist}(\phi(\rho), \phi'(\rho')) \vee \mathrm{dist}_{\mathrm{H}}(\phi(X), \phi'(X'))$$

where ϕ, ϕ' are isometries from X, X' into a common metric space (M, dist), and where $\mathrm{dist}_{\mathrm{H}}$ is the Hausdorff distance between compact subsets of (M, dist). It is elementary that this distance does not depend on particular choices in the equivalence classes of (X, ρ) and (X', ρ'). We endow $\mathscr{M}$ with the associated Borel σ-algebra. Of course, d_{GH} satisfies a homogeneity property, $d_{\mathrm{GH}}(aX, aX') = a d_{\mathrm{GH}}(X, X')$ for $a > 0$.

We also need to define a distance on $\mathscr{M}_{\mathrm{w}}$, that is in some sense compatible with the Gromov-Hausdorff distance. Several complete distances can be constructed, and we use a variation of the Gromov-Hausdorff-Prokhorov distance used in [30]. The induced topology is the same as that introduced earlier in [22]. The reader should bear in mind that the topology used in the present paper involves a little extension of the two previous references, since we are interested in pointed spaces. We let $d_{\mathrm{GHP}}((X, d, \rho, \mu), (X', d', \rho', \mu'))$ be the infimum of the quantities

$$\mathrm{dist}(\phi(\rho), \phi'(\rho')) \vee \mathrm{dist}_{\mathrm{H}}(\phi(X), \phi'(X')) \vee \mathrm{dist}_{\mathrm{P}}(\phi_* \mu, \phi'_* \mu') \, ,$$

where again ϕ, ϕ' are isometries from X, X' into a common space (M, dist), $\phi_*\mu, \phi'_*\mu'$ are the push-forward of μ, μ' by ϕ, ϕ', and $\mathrm{dist}_{\mathrm{P}}$ is the Prokhorov distance between Borel probability measures on M [20, Chapter 3]:

$$\mathrm{dist}_{\mathrm{P}}(m, m') = \inf\{\varepsilon > 0 : m(C) \leq m'(C^\varepsilon) + \varepsilon \text{ for every } C \subset M \text{ closed}\} \, ,$$

where $C^\varepsilon = \{x \in M : \inf_{y \in C} \mathrm{dist}(x, y) < \varepsilon\}$ is the ε-thickening of C. A simple adaptation of the results of [22] and Section 6 in [30] (in order to take into account the particular role of the distinguished point ρ) shows that:

Proposition 1. *The function d_{GHP} is a distance on $\mathscr{M}_{\mathrm{w}}$ that makes it complete and separable.*

This distance is called the pointed Gromov-Hausdorff-Prokhorov distance. One must be careful that contrary to d_{GH}, this distance is not homogeneous: $d_{\mathrm{GHP}}(aX, aX')$ is in general different from $a d_{\mathrm{GHP}}(X, X')$, because only the distances, not the measures, are multiplied in aX, aX'.

[1] usually such spaces are rather called *pointed*, but we prefer the term rooted which is more common when dealing with trees

1.3.1 Trees viewed as metric spaces

A plane tree $\mathfrak{t}$ can be naturally seen as a metric space by endowing $\mathfrak{t}$ with the *graph distance* between vertices. Namely,

$$d_{\mathrm{gr}}(u,v) = |u| + |v| - 2|u \wedge v|\,, \qquad u, v \in \mathfrak{t}\,,$$

where $u \wedge v$ is the longest prefix common to u, v. This coincides with the number of edges on the only simple path going from u to v. The space $(\mathfrak{t}, d_{\mathrm{gr}})$ is naturally rooted at $\varnothing$. We can put two natural probability measures on $\mathfrak{t}$, the uniform measures on the leaves or on the vertices:

$$\mu_{\partial\mathfrak{t}} = \frac{1}{\#\partial\mathfrak{t}} \sum_{u \in \partial\mathfrak{t}} \delta_{\{u\}}\,, \qquad \mu_{\mathfrak{t}} = \frac{1}{\#\mathfrak{t}} \sum_{u \in \mathfrak{t}} \delta_{\{u\}}\,.$$

If $\mathsf{t} \in \mathsf{T}$ is a tree and $\mathfrak{t}, \mathfrak{t}'$ are two plane representatives of t, then it is elementary that the spaces $(\mathfrak{t}, d_{\mathrm{gr}}, \varnothing, \mu_{\partial\mathfrak{t}})$ and $(\mathfrak{t}', d_{\mathrm{gr}}, \varnothing, \mu_{\partial\mathfrak{t}'})$ are isometry-equivalent rooted measured metric spaces. The same holds with $\mu_{\mathfrak{t}}, \mu_{\mathfrak{t}'}$ instead of $\mu_{\partial\mathfrak{t}}, \mu_{\partial\mathfrak{t}'}$. We denote by $(\mathsf{t}, d_{\mathrm{gr}}, \rho, \mu_{\partial\mathsf{t}})$ and $(\mathsf{t}, d_{\mathrm{gr}}, \rho, \mu_{\mathsf{t}})$ the corresponding elements of $\mathscr{M}_{\mathrm{w}}$. Conversely, it is possible to recover uniquely the discrete tree (not a plane tree!) from the element of $\mathscr{M}_{\mathrm{w}}$ thus defined.

1.3.2 $\mathbb{R}$-trees

An $\mathbb{R}$-tree is a metric space (X, d) such that for every $x, y \in X$,

1. there is an isometry $\varphi_{x,y} : [0, d(x,y)] \to X$ such that $\varphi_{x,y}(0) = x$ and $\varphi_{x,y}(d(x,y)) = y$, and
2. for every continuous, injective function $c : [0,1] \to X$ with $c(0) = x$, $c(1) = y$, one has $c([0,1]) = \varphi_{x,y}([0, d(x,y)])$.

In other words, any two points in X are linked by a geodesic path, which is the only simple path linking these points, up to reparameterisation. This is a continuous analogue of the graph-theoretic definition of a tree as a connected graph with no cycle. We denote by $[[x,y]]$ the range of $\varphi_{x,y}$.

We let $\mathscr{T}$ (resp. $\mathscr{T}_{\mathrm{w}}$) be the set of isometry classes of compact rooted $\mathbb{R}$-trees (resp. compact, rooted measured $\mathbb{R}$-trees). An important property is the following (these are easy variations on results by [21, 22]).

Proposition 2. *The spaces $\mathscr{T}$ and $\mathscr{T}_{\mathrm{w}}$ are closed subspaces of $(\mathscr{M}, d_{\mathrm{GH}})$ and $(\mathscr{M}_{\mathrm{w}}, d_{\mathrm{GHP}})$.*

If $\mathcal{T} \in \mathscr{T}$ and for $x \in \mathcal{T}$, we call the quantity $d(\rho, x)$ the *height* of x. If $x, y \in \mathcal{T}$, we say that x is an ancestor of y whenever $x \in [[\rho, y]]$. We let $x \wedge y \in \mathcal{T}$ be the unique element of $\mathcal{T}$ such that $[[\rho, x]] \cap [[\rho, y]] = [[\rho, x \wedge y]]$, and call it the highest common ancestor of x and y in $\mathcal{T}$. For $x \in \mathcal{T}$, we denote by $\mathcal{T}_x$ the set of $y \in \mathcal{T}$ such that x is an ancestor of y. The set $\mathcal{T}_x$, endowed with the restriction of the distance d, and rooted at x, is in turn a rooted $\mathbb{R}$-tree, called the subtree of $\mathcal{T}$ rooted at x. If $(\mathcal{T}, d, \rho, \mu)$ is an element of $\mathscr{T}_{\mathrm{w}}$ and $\mu(\mathcal{T}_x) > 0$, then $\mathcal{T}_x$ can be seen as an element of $\mathscr{T}_{\mathrm{w}}$ by endowing it with the measure $\mu(\cdot \cap \mathcal{T}_x)/\mu(\mathcal{T}_x)$.

We say that $x \in \mathcal{T}$, $x \neq \rho$, in a rooted $\mathbb{R}$-tree is a *leaf* if its removal does not disconnect $\mathcal{T}$. Note that this always exclude the root from the set of leaves, which we denote by $\mathcal{L}(\mathcal{T})$. A *branch point* is an element of $\mathcal{T}$ of the form $x \wedge y$ where x is not an ancestor of y nor vice-versa. It is also characterized by the fact that the removal of a branch point disconnects the $\mathbb{R}$-tree into three or more components (two or more for the root). We let $\mathcal{B}(\mathcal{T})$ be the set of branch points of $\mathcal{T}$.

1.4 Self-similar fragmentations and associated $\mathbb{R}$-trees

Self-similar fragmentation processes are continuous-time processes that describe the dislocation of a massive object as time passes. Introduce the set of partitions of a unit mass

$$\mathcal{S}^{\downarrow} := \left\{ \mathbf{s} = (s_1, s_2, \ldots) : s_1 \geq s_2 \geq \ldots \geq 0, \sum_{i \geq 1} s_i \leq 1 \right\}.$$

This space is endowed with the ℓ^1-metric $d(\mathbf{s}, \mathbf{s}') = \sum_{i \geq 1} |s_i - s'_i|$, which makes it a compact space.

Definition 1. *A self-similar fragmentation is an $\mathcal{S}^{\downarrow}$-valued Markov process $(\mathbf{X}(t), t \geq 0)$ which is continuous in probability and satisfies the following fragmentation property. For some $a \in \mathbb{R}$, called the self-similarity index, it holds that conditionally given $\mathbf{X}(t) = (s_1, s_2, \ldots)$, the process $(\mathbf{X}(t + t'), t' \geq 0)$ has same distribution as the process whose value at time t' is the decreasing rearrangement of the sequences $s_i \mathbf{X}^{(i)}(s_i^a t'), i \geq 1$, where $(\mathbf{X}^{(i)}, i \geq 1)$ are i.i.d. copies of $\mathbf{X}$.*

Bertoin [8] and Berestycki [6] have shown that the laws of self-similar fragmentation processes are characterized by three parameters: the index a, a non-negative erosion coefficient, and a dislocation measure ν on $\mathcal{S}^{\downarrow}$. The idea is that every sub-object of the initial object, with mass x say, will suddenly split into sub-sub-objects of masses $xs_1, xs_2, \ldots$ at rate $x^a \nu(\mathrm{d}\mathbf{s})$, independently of the other sub-objects. Erosion accounts for the formation of zero-mass particles that are continuously ripped off the fragments.

For our concerns, we will consider only the special case where the erosion phenomenon has no role and the dislocation measure does not charge the set $\{\mathbf{s} \in \mathcal{S}^{\downarrow} : \sum_i s_i < 1\}$. One says that ν is *conservative.* This motivates the following definition.

Definition 2. *A* dislocation measure *is a σ-finite measure ν on $\mathcal{S}^{\downarrow}$ such that $\nu(\{(1, 0, 0, \ldots)\}) = 0$ and*

$$\nu\Big(\Big\{ \sum_{i \geq 1} s_i < 1 \Big\}\Big) = 0, \qquad \int_{\mathcal{S}^{\downarrow}} (1 - s_1) \nu(\mathrm{d}\mathbf{s}) < \infty. \tag{2}$$

We say that the measure is binary *when $\nu(\{s_1 + s_2 < 1\}) = 0$. A binary measure is characterized by its image $\nu(s_1 \in \mathrm{d}x)$ through the mapping $\mathbf{s} \mapsto s_1$.*

A fragmentation pair *is a pair (a, ν) where $a \in \mathbb{R}$ is called the self-similarity index, and ν is a dislocation measure.*

Fragmentation pairs (a, ν) therefore characterize the distributions of the self-similar fragmentations we are focusing on. When $a = -\gamma < 0$, small fragments tend to split faster, and it turns out that they all disappear in finite time, a property known as *formation of dust.* Using this property, it is shown in [24] how to construct a "continuum random tree" encoding the process with characteristic pair $(-\gamma, \nu)$. This tree is a random element of $\mathscr{M}_{\mathrm{w}}$,

$$\mathcal{T}_{\gamma,\nu} = (\mathcal{T}, d, \rho, \mu),$$

such that $(\mathcal{T}, d)$ is a random $\mathbb{R}$-tree, and such that almost-surely

1. the measure μ is supported on the set $\mathcal{L}(\mathcal{T})$ of leaves of $\mathcal{T}$
2. μ has no atom,
3. for every $x \in \mathcal{T} \setminus \mathcal{L}(\mathcal{T})$, it holds that $\mu(\mathcal{T}_x) > 0$.

Proposition 3 (Theorem 1 and Proposition 1 in [24])**.** *The law of $\mathcal{T}_{\gamma,\nu}$ is characterized by properties 1, 2, 3 above, together with the fact that if $t \geq 0$ and $\mathcal{T}_i(t), i \geq 1$ are the connected components of the open set $\{x \in \mathcal{T} : d(\rho, x) > t\}$, then the process $((\mu(\mathcal{T}_i(t)), i \geq 1)^{\downarrow}, t \geq 0)$ of the non-increasing rearrangement of the μ-masses of these components has the same law as the $\mathcal{S}^{\downarrow}$-valued self-similar fragmentation with characteristic pair $(-\gamma, \nu)$.*

The previous proposition is not really constructive, and we postpone a more detailed description of $\mathcal{T}_{\gamma,\nu}$ to Section 3.2.

It was shown in [24] that one can recover the celebrated Brownian and stable continuum random trees [1, 28, 18] as special instances of fragmentation trees. The parameters γ and ν corresponding to these trees will be recalled when we discuss applications in Sections 2.1 and 2.2.

1.5 Main results

Let $(q_n(\lambda), \lambda \in \mathcal{P}_n), n \geq 1$ satisfy (1). With it, we associate a finite non-negative measure $\overline{q}_n$ on $\mathcal{S}^{\downarrow}$, defined by its integral against measurable $f : \mathcal{S}^{\downarrow} \to \mathbb{R}_+$ as

$$\overline{q}_n(f) = \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) f\left(\frac{\lambda}{n}\right).$$

Note that in the left-hand side, we have identified λ/n with an element of $\mathcal{S}^{\downarrow}$, in accordance with our convention that λ is identified with the infinite sequence $(\lambda_i, i \geq 1)$. We make the following basic assumption.

> **(H)** There exists a fragmentation pair $(-\gamma, \nu)$, with $\gamma > 0$, and a function $\ell : (0, \infty) \to (0, \infty)$ slowly varying at ∞, such that we have the weak convergence of finite non-negative measures on $\mathcal{S}^{\downarrow}$:
>
> $$n^{\gamma} \ell(n)(1 - s_1)\overline{q}_n(\mathrm{d}\mathbf{s}) \xrightarrow[n\to\infty]{(w)} (1 - s_1)\nu(\mathrm{d}\mathbf{s}). \tag{3}$$

Theorem 1. *Assume $q = (q_n(\lambda), \lambda \in \mathcal{P}_n), n \geq 1$ satisfies assumption* **(H)***. Let T_n have distribution P^q_n, and view T_n as a random element of $\mathscr{M}_{\mathrm{w}}$ by endowing it with the graph distance and the uniform probability measure $\mu_{\partial T_n}$ on ∂T_n. Then we have the convergence in distribution*

$$\frac{1}{n^{\gamma}\ell(n)} T_n \xrightarrow[n\to\infty]{(d)} \mathcal{T}_{\gamma,\nu},$$

for the rooted Gromov-Hausdorff-Prokhorov topology.

There is a similar statement for the trees with laws Q^q_n. Consider a family $(q_n(\lambda), \lambda \in \mathcal{P}_n), n \geq 1$ with $q_1((1)) = 1$.

Theorem 2. *Assume $q = (q_n(\lambda), \lambda \in \mathcal{P}_n), n \geq 1$ satisfies assumption* **(H)***, with*

- *either $\gamma \in (0, 1)$, or*
- *$\gamma = 1$ and $\ell(n) \to 0$ as $n \to \infty$.*

Let T_n have distribution Q^q_n. We view T_n as a random element of $\mathscr{M}_{\mathrm{w}}$ by endowing it with the graph distance and the uniform probability measure μ_{T_n} on T_n. Then we have the convergence in distribution

$$\frac{1}{n^{\gamma}\ell(n)} T_n \xrightarrow[n\to\infty]{(d)} \mathcal{T}_{\gamma,\nu},$$

for the rooted Gromov-Hausdorff-Prokhorov topology.

Theorem 2 deals with a more restricted set of values of values of γ than Theorem 1. This comes from the fact that, contrary to the set T^{∂}_n which contains trees with arbitrary height, the set T_n of trees with n vertices has elements with height at most $n - 1$. Therefore, we cannot hope to find non-trivial limits in Theorem 2 when $\gamma > 1$, or when $\gamma = 1$ and $\ell(n)$ has limit $+\infty$ as $n \to \infty$. The intermediate case where $\ell(n)$ admits finite non-zero limiting points cannot give such a convergence

with a continuum fragmentation tree in the limit either. The reason being that the support of the height of a continuum fragmentation tree is unbounded, whereas the heights of $T_n/n\ell(n)$ are all bounded from above by $1/\inf_n(\ell(n))$, which is finite under our assumption.

Note that Theorem 1 (resp. Theorem 2) implies that any fragmentation tree $\mathcal{T}_{\gamma,\nu}$ is the continuous limit of a rescaled family of discrete Markov branching trees with a prescribed number of leaves (resp. with a prescribed number of vertices, provided $\gamma < 1$), since we have the following approximation result.

Proposition 4. *For every fragmentation pair $(-\gamma, \nu)$ with $\gamma > 0$, there exists a family of distributions $(q_n, n \geq 1)$ satisfying (1) and such that (3) holds, with $\ell(x) = 1$ for every $x > 0$.*

After some preliminaries gathered in Section 3, we prove Theorems 1 and 2 and Proposition 4 in Section 4. Before embarking in the proofs, we present in Section 2 some important applications of these theorems to Galton-Watson trees, unordered random trees and particular families of Markov branching trees studied in earlier works. Of these applications, the first two actually involve a substantial amount of work, so that the details are postponed to Section 5 and 6.

2 Applications

2.1 Galton-Watson trees

A natural application is the study of Galton-Watson trees conditioned on their total number of vertices. Let ξ be a probability measure on $\mathbb{Z}_+$ such that $\xi(0) > 0$ and

$$\sum_{k\geq 0} k\xi(k) = 1\,. \tag{4}$$

The law of the Galton-Watson tree with offspring distribution ξ is the probability measure on the set of plane trees defined by

$$\mathrm{GW}_\xi(\{\mathbb{t}\}) = \prod_{u\in\mathbb{t}} \xi(c_u(\mathbb{t}))\,,$$

for $\mathbb{t}$ a plane tree. That this does define a probability distribution on the set of plane trees comes from the fact that a Galton-Watson process with offspring distribution ξ becomes a.s. extinct in finite time, due to the criticality condition (4). In order to fit in the framework of this paper, we view GW_ξ as a distribution on the set of discrete, rooted trees, by taking its push-forward under the natural projection from plane trees to trees.

In order to avoid technicalities, we also assume that the support of ξ generates the additive group $\mathbb{Z}$. This implies that $\mathrm{GW}_\xi(\{\#\mathbb{t} = n\}) > 0$ for every n large enough. For such n, we let $\mathrm{GW}_\xi^{(n)} = \mathrm{GW}_\xi(\cdot \,|\, \{\#\mathbb{t} = n\})$, and view it as a law on T_n.

We distinguish two different regimes.

Case 1. The offspring distribution has finite variance

$$\sigma^2 = \sum_{k\geq 0} k(k-1)\xi(k) < \infty\,.$$

Case 2. For some $\alpha \in (1, 2)$ and $c \in (0, \infty)$, it holds that $\xi(k) \sim ck^{-\alpha-1}$ as $k \to \infty$. In particular, ξ is in the domain of attraction of a stable law of index α.

The *Brownian dislocation measure* is the unique binary dislocation measure such that

$$\nu_2(s_1 \in \mathrm{d}x) = \sqrt{\frac{2}{\pi x^3(1-x)^3}}\,\mathrm{d}x\,\mathbf{1}_{\{1/2\leq x<1\}}\,.$$

Otherwise said, for every measurable $f : \mathcal{S}^{\downarrow} \to \mathbb{R}_+$,

$$\int_{\mathcal{S}^{\downarrow}} \nu_2(\mathrm{d}\mathbf{s})\, f(\mathbf{s}) = \int_{1/2}^{1} \sqrt{\frac{2}{\pi x^3(1-x)^3}}\, \mathrm{d}x\, f(x, 1-x, 0, 0, \ldots)\,.$$

We also define a one-parameter family of measures in the following way. For $\alpha \in (1,2)$, let $\sum_{i\geq 1} \delta_{\Delta_i}$ be a Poisson random measure on $(0,\infty)$ with intensity measure

$$\frac{1}{\alpha\Gamma\big(1-\frac{1}{\alpha}\big)} \frac{\mathrm{d}x}{x^{1+1/\alpha}} \mathbf{1}_{\{x>0\}}\,,$$

with the atoms $\Delta_i, i \geq 1$ labeled in such a way that $\Delta_1 \geq \Delta_2 \geq \ldots$. Let $T = \sum_{i\geq 1} \Delta_i$, which is finite a.s. by standard properties of Poisson measures. In fact, T follows a stable distribution with index $1/\alpha$, with Laplace transform

$$\mathbb{E}[\exp(-\lambda T)] = \exp(-\lambda^{1/\alpha})\,, \qquad \lambda \geq 0\,.$$

It can be seen as a stable subordinator evaluated at time 1, its jumps up to this time being the atoms $\Delta_i, i \geq 1)$. The measure ν_α is defined by its action against a measurable function $f : \mathcal{S}^{\downarrow} \to \mathbb{R}_+$:

$$\int_{\mathcal{S}^{\downarrow}} \nu_\alpha(\mathrm{d}\mathbf{s})\, f(\mathbf{s}) = \frac{\alpha^2\Gamma(2-1/\alpha)}{\Gamma(2-\alpha)} \mathbb{E}\left[T\, f\Big(\frac{\Delta_i}{T}, i \geq 1\Big)\right].$$

Because $\mathbb{E}[T] = \infty$, this formula defines an infinite σ-finite measure on $\mathcal{S}^{\downarrow}$, which turns out to satisfy (2).

Theorem 3. *Let ξ satisfy (4), with support that generates the additive group $\mathbb{Z}$. Let T_n be a random element of T_n with distribution $\mathrm{GW}_\xi^{(n)}$. Consider T_n as an element of $\mathscr{M}_{\mathrm{w}}$ by endowing it with the graph distance and the uniform probability measure μ_{T_n} on T_n. Then we have, in distribution for the Gromov-Hausdorff-Prokhorov topology,*

$$\textbf{Case 1}: \qquad \frac{1}{\sqrt{n}} T_n \xrightarrow[n\to\infty]{(d)} \frac{2}{\sigma} \mathcal{T}_{1/2,\nu_2}\,,$$

$$\textbf{Case 2}: \qquad \frac{1}{n^{1-1/\alpha}} T_n \xrightarrow[n\to\infty]{(d)} \left(\frac{\alpha(\alpha-1)}{c\Gamma(2-\alpha)}\right)^{1/\alpha} \mathcal{T}_{1-1/\alpha,\nu_\alpha}\,.$$

This result will be proved in Section 5 below, by first showing that $\mathrm{GW}_\xi^{(n)}$ is of the form Q_n^q for some appropriate choice of q.

The trees $\mathcal{T}_{1/2,\nu_2}$ and $\mathcal{T}_{1-1/\alpha,\nu_\alpha}$ appearing in the limit are important models of continuum random trees, called respectively the Brownian Continuum Random Tree and the stable tree with index α. The Brownian tree is somehow the archetype in the theory of scaling limits of trees. The above theorem is very similar to a result due to Duquesne [17], but our method of proof is totally different. While [17] relies on quite refined aspects of Galton-Watson trees and their encodings by stochastic processes, our approach requires only to have some kind of global structure, namely the Markov branching property, and to know how mass is spread in one generation. We do not claim that our method is more powerful than the one used in [17] (as a matter of fact, the limit theorem of [17] holds in the more general case where μ is in the domain of attraction of a totally asymmetric stable law with index $\alpha \in (1,2]$). However, our method has some robustness, allowing to shift from Galton-Watson trees to other models of trees. Our next example will try to illustrate this.

2.2 Uniform unordered trees

Our next application is on a different model of random trees, which is by nature not a model of plane or labelled trees, contrary to the previous examples. It is actually not either a Markov branching model, but is very close from being one, as we will see.

For $2 \leq m \leq \infty$, we consider the set $\mathsf{T}_n^{(m)} \subset \mathsf{T}_n$ of trees with n vertices, in which every vertex has at most m children. In particular, we have $\mathsf{T}_n^{(\infty)} = \mathsf{T}_n$. The sets $\mathsf{T}_n^{(m)}$ are harder to enumerate than sets of ordered or labeled trees, like plane trees or Cayley trees, and there is no closed expression for the numbers $\#\mathsf{T}_n^{(m)}$. However, Otter [31] (see also [23, Section VII.5]) derived the asymptotic enumeration result

$$\#\mathsf{T}_n^{(m)} \underset{n\to\infty}{\sim} \kappa_m \frac{(\rho_m)^n}{n^{3/2}}\,, \tag{5}$$

for some m-dependent constants $\kappa_m > 0, \rho_m > 1$. This can be achieved by studying the generating function

$$\psi^{(m)}(x) = \sum_{n\geq 1} \#\mathsf{T}_n^{(m)}\, x^n\,,$$

which has a square-root singularity at the point $1/\rho_m$. The behavior (5) indicates that a uniformly chosen element of $\mathsf{T}_n^{(m)}$ should converge as $n \to \infty$, once renormalized suitably, to the Brownian continuum random tree. We show that this is indeed the case for any value of m. To state our result, let

$$\widetilde{\mathsf{T}}_n^{(m)} = \{\mathsf{t} \in \mathsf{T}_n^{(m)} : c_\varnothing(\mathsf{t}) \leq m-2\}\,.$$

For instance, $\widetilde{\mathsf{T}}_n^{(2)} = \emptyset$ for $n \geq 2$, while $\widetilde{\mathsf{T}}_n^{(\infty)} = \mathsf{T}_n^{(\infty)}$ for all n. Let

$$\widetilde{\psi}^{(m)}(x) = \sum_{n\geq 1} \#\widetilde{\mathsf{T}}_n^{(m)}\, x^n\,,$$

and define a finite constant c_m by

$$c_m = \frac{\sqrt{2}}{\sqrt{\pi}\kappa_m \widetilde{\psi}^{(m)}(1/\rho_m)}\,.$$

Note that $\widetilde{\psi}^{(2)}(x) = x$ for every x, while $\widetilde{\psi}^{(\infty)}(1/\rho_\infty) = 1$ [23, Section VII.5]. Therefore, we get

$$c_2 = \frac{\sqrt{2}\rho_2}{\sqrt{\pi}\kappa_2}\,, \qquad c_\infty = \frac{\sqrt{2}}{\sqrt{\pi}\kappa_\infty}\,.$$

Theorem 4. *Fix $m \in \{2, 3, \dots\} \cup \{\infty\}$. Let T_n be uniformly distributed in $\mathsf{T}_n^{(m)}$. We view T_n as an element of $\mathscr{M}_{\mathrm{w}}$ by endowing it with the measure μ_{T_n}, then*

$$\frac{1}{\sqrt{n}} T_n \xrightarrow[n\to\infty]{(d)} c_m \mathcal{T}_{1/2,\nu_2}\,.$$

The proof of this result is given in Section 6. We note that this implies a similar, maybe more natural, statement for m-ary trees. We say that $\mathsf{t} \in \mathsf{T}$ is m-ary if every vertex has either m children or no child, and we say that the vertex is internal in the first case, i.e. when it is not a leaf. Summing over the degrees of vertices in an m-ary tree with n internal vertices, we obtain that $\#\mathsf{t} = mn + 1$, and $\#\partial\mathsf{t} = (m-1)n + 1$.

Assume now that $m < \infty$. Starting from an m-ary tree t with n internal vertices, and removing the leaves — equivalently, keeping only the internal vertices — gives an element $\phi(\mathsf{t}) \in \mathsf{T}_n^{(m)}$. The mapping ϕ is inverted by attaching $m-k$ leaves to each vertex with k children, for an element of $\mathsf{T}_n^{(m)}$. Moreover, we leave as an easy exercise that $d_{\mathrm{GHP}}(a\mathsf{t}, a\phi(\mathsf{t})) \leq a$ for every $a > 0$, when the trees are endowed with the uniform measures $\mu_{\mathsf{t}}, \mu_{\phi(\mathsf{t})}$ on vertices. Theorem 4 thus implies:

Corollary 1. *Let $m \in \{2, 3, \dots,\}$ and $T_n^{[m]}$ be a uniform m-ary tree with n internal vertices, endowed with the measure $\mu_{T_n^{[m]}}$. Then*

$$\frac{1}{\sqrt{n}} T_n^{[m]} \xrightarrow[n\to\infty]{(d)} c_m \mathcal{T}_{1/2,\nu_2} .$$

The problem of scaling limits of random rooted unlabeled trees has attracted some attention in the recent literature, see [13, 29, 16]. For $m = 2$, Corollary 1 readily yields the main theorem of [29], which was derived using a completely different method. Indeed, it is based in a stronger way on combinatorial aspects of $\mathsf{T}_n^{(2)}$. Here, we really make use of a fragmentation property satisfied by the laws $\mathsf{P}_n^{(m)}$. As alluded to at the begining of this section, these are not actually laws of Markov branching trees. Nevertheless, they can be coupled with laws of Markov branching trees in a way that the coupled trees are close in the d_{GHP} metric. In the general case $m \neq 2$, Theorem 4 and Corollary 1 are new, and were implicitly conjectured by Aldous [2]. In [16], the authors prove a result on the scaling limit of the so-called profile of the uniform tree for $m = \infty$, which is related to our results, although it is not a direct consequence. Finally, we note that the problem of the scaling limit of *unrooted* unordered trees is still open, although we expect the Brownian tree to arise again as the limiting object.

2.3 Consistent Markov branching models

Considering again in a more specific way the Markov branching models, we stress that Theorem 1 also encompasses the results of [26], which hold for particular families $(q_n, n \geq 1)$ satisfying a further consistency property. In this setting, it is assumed that $q_n((n)) = 0$ for every $n \geq 1$, so that the trees T_n do not have any vertex having only one child. The consistency property can be formulated as follows:

> **Consistency property.** Starting from T_n with $n \geq 2$, select one of the leaves uniformly at random, and remove this leaf as well as the edge that is attached to it. If this removal creates a vertex with only one child, then remove this vertex and merge the two edges incident to this vertex into one. Then the random tree thus constructed has same distribution as T_{n-1}.

A complete characterization of families $(q_n, n \geq 1)$ giving rise to Markov branching trees with this consistency property is given in [26]. Namely, such families can be constructed in terms of a pair (c, ν), which is uniquely defined up to multiplication by a common positive constant, such that $c \geq 0$ is an "erosion coefficient" and ν is a dislocation measure as described above (except that $\nu(\sum s_i < 1) = 0$ is not required). The cases where $c = 0$ and $\nu(\sum s_i < 1) = 0$ are the most interesting ones, so we will assume henceforth that this is the case. The associated distributions $q_n, n \geq 2$ are given by the following explicit formula: for $\lambda \in \mathcal{P}_n$ having $p \geq 2$ parts,

$$q_n(\lambda) = \frac{1}{Z_n} C_\lambda \int_{\mathcal{S}^\downarrow} \nu(\mathrm{d}\mathbf{s}) \sum_{\substack{i_1,\dots,i_p \geq 1 \\ \text{distinct}}} \prod_{j=1}^{p} s_{i_j}^{\lambda_j} , \tag{6}$$

where

$$C_\lambda = \frac{n!}{\prod_{i\geq 1} \lambda_i! \prod_{j\geq 1} m_j(\lambda)!}$$

is a combinatorial factor, the same that appears in the statement of Lemma 5 below, and Z_n is a normalizing constant defined by

$$Z_n = \int_{\mathcal{S}^\downarrow} \nu(\mathrm{d}\mathbf{s}) \Big(1 - \sum_{i\geq 1} s_i^n \Big) .$$

Assume further that ν satisfies the following regularity condition:

$$\nu(s_1 \leq 1-\varepsilon) = \varepsilon^{-\gamma}\ell(1/\varepsilon)\,, \tag{7}$$

where $\gamma \in (0,1)$ and ℓ is a function that is slowly varying at ∞. Then

Theorem 5. *If ν is a dislocation measure satisfying (2) and (7), and if $(q_n, n \geq 1)$ is the consistent family of probability measures defined by (6), then the Markov branching trees T_n, viewed as random measured $\mathbb{R}$-trees by endowing the set of their leaves with the uniform probability measures, satisfies*

$$\frac{1}{\Gamma(1-\gamma)n^{\gamma}\ell(n)} T_n \xrightarrow[n\to\infty]{(d)} \mathcal{T}_{\gamma_\nu,\nu}\,,$$

for the Gromov-Hausdorff-Prokhorov topology.

This theorem is in some sense more powerful than [26, Theorem 2], because the latter result needed one extra technical hypothesis that is discarded here. Moreover, our result holds for the Gromov-Hausdorff-Prokhorov topology, which is stronger than the Gromov-Hausdorff topology considered in [26]. However, the setting of [26] also provided a natural coupling of the trees $T_n, n \geq 1$ and $\mathcal{T}_{\gamma_\nu,\nu}$ on the same probability space, for which the convergence in Theorem 5 can be strengthened to a convergence in probability. This coupling is not provided in our case.

Proof. Let $\mathbf{s} \in \mathcal{S}^{\downarrow}$ be such that $\sum_{i\geq 1} s_i = 1$. Let $K_1, \ldots, K_n$ be i.i.d. random variables in $\mathbb{N}$ such that $\mathbb{P}(K_1 = i) = s_i$ for every $i \geq 1$. Call $\Lambda^{(i)}(n)$ the number of variables K_j equal to i, and let $\Lambda^{(\mathbf{s})}(n) = \left(\Lambda^{(1)}(n), \Lambda^{(2)}(n), \ldots\right)^{\downarrow}$, where $\mathbf{x}^{\downarrow}$ denotes the decreasing rearrangement of the non-negative sequence $\mathbf{x} = (x_1, x_2, \ldots)$ with finite sum. It is not hard to see that the probability distributions q_n defined by (6) are also given, for $\lambda \neq (n)$, by

$$q_n(\lambda) = \frac{1}{Z_n}\int_{\mathcal{S}^{\downarrow}} \mathbb{P}(\Lambda^{(\mathbf{s})}(n) = \lambda)\nu(\mathrm{d}\mathbf{s}).$$

See for example the forthcoming Lemma 5 in Section 3.2.4. The normalizing constant Z_n is regularly varying, according to the assumption of regular variation (7). Indeed, by Karamata's Tauberian Theorem (see [11, Th. 1.7.1']), we have that

$$Z_n = \int_{\mathcal{S}^{\downarrow}} \left(1 - \sum_{i\geq 1} s_i^n\right)\nu(\mathrm{d}\mathbf{s}) \underset{n\to\infty}{\sim} \int_{\mathcal{S}^{\downarrow}} (1-s_1^n)\nu(\mathrm{d}\mathbf{s}) \underset{n\to\infty}{\sim} \Gamma(1-\gamma)\nu(s_1 \leq 1-1/n) = \Gamma(1-\gamma)n^{\gamma}\ell(n).$$

Now, to get a convergence of the form (3), note that for all continuous functions $f : \mathcal{S}^{\downarrow} \to \mathbb{R}_+$,

$$\begin{aligned} Z_n \sum_{\lambda\in\mathcal{P}_n} q_n(\lambda)\left(1-\frac{\lambda_1}{n}\right) f\left(\frac{\lambda}{n}\right) \quad &= \quad \int_{\mathcal{S}^{\downarrow}} \nu(\mathrm{d}\mathbf{s})\mathbb{E}\left[\left(1-\frac{\Lambda_1^{(\mathbf{s})}(n)}{n}\right) f\left(\frac{\Lambda^{(\mathbf{s})}(n)}{n}\right)\right] \\ &\underset{n\to\infty}{\to} \quad \int_{\mathcal{S}^{\downarrow}} \nu(\mathrm{d}\mathbf{s})(1-s_1)f(\mathbf{s})\,, \end{aligned}$$

which follows by dominated convergence, since f is bounded (say by K) on the compact space $\mathcal{S}^{\downarrow}$ and

$$\mathbb{E}\left[\left(1-\frac{\Lambda_1^{(\mathbf{s})}(n)}{n}\right) f\left(\frac{\Lambda^{(\mathbf{s})}(n)}{n}\right)\right] \leq K\mathbb{E}\left[1-\frac{\Lambda_1^{(\mathbf{s})}(n)}{n}\right] \leq K\mathbb{E}\left[1-\frac{\Lambda^{(1)}(n)}{n}\right] = K(1-s_1)\,.$$

We conclude by applying Theorem 1. $\square$

2.4 Further non-consistent cases: (α,θ)-trees

Another application concerns a family of binary labeled trees introduced by Pitman and Winkel [33] and built inductively according to a growth rule depending on two parameters $\alpha \in (0,1)$ and $\theta \geq 0$. Roughly, at each step, given that the tree $T_n^{\alpha,\theta,\mathrm{lab}}$ with n leaves branches at the branch point adjacent to the root into two subtrees with $k \geq 1$ leaves for the subtree containing the smallest label in $T_n^{\alpha,\theta,\mathrm{lab}}$ and $n-k \geq 1$ leaves for the other one, a weight α is assigned to the root edge and weights $k-\alpha$ and $n-k-1+\theta$ are assigned respectively to the trees with sizes k, $n-k$. Then choose either the root edge or one of the two subtrees according with probabilities proportional to these weights. If a subtree with two or more leaves is selected, apply this weighting procedure inductively to this subtree until the root edge or a subtree with a single leaf is selected. If a subtree with single leaf is selected, insert a new edge and leaf at the unique edge of this subtree. Similarly, if the root edge is selected, add a new edge and leaf to this root edge. We denote by $T_n^{\alpha,\theta}$ the tree $T_n^{\alpha,\theta,\mathrm{lab}}$ without labels.

Pitman and Winkel show that the family $(T_n^{\alpha,\theta}, n \geq 1)$ is not consistent in general ([33, Proposition 1]), except when $\theta = 1-\alpha$ or $\theta = 2-\alpha$, and has the Markov branching property ([33, Proposition 11]) with the following probabilities q_n:

- $q_n\left((k, n-k, 0, \ldots)\right) = q_{\alpha,\theta}(n-1,k) + q_{\alpha,\theta}(n-1,n-k), \quad$ for $n-k < k \leq n-1$
- $q_n\left(n/2, n/2\right) = q_{\alpha,\theta}(n-1,n/2)$,

where

$$q_{\alpha,\theta}(n,k) = \binom{n}{k} \frac{\alpha(n-k)+\theta k}{n} \frac{\Gamma(k-\alpha)\Gamma(n-k+\theta)}{\Gamma(1-\alpha)\Gamma(n+\theta)}, \quad 1 \leq k \leq n.$$

Now consider the binary measure $\nu_{\alpha,\theta}$ defined on $\mathcal{S}^{\downarrow}$ by $\nu_{\alpha,\theta}(s_1+s_2<1)=0$ and $\nu_{\alpha,\theta}(s_1 \in \mathrm{d}x) = f_{\alpha,\theta}(x)\mathrm{d}x$ where $f_{\alpha,\theta}$ is defined on $[1/2,1)$ by

$$f_{\alpha,\theta}(x) = \frac{1}{\Gamma(1-\alpha)}\left(\left(\alpha\,(1-x)+\theta x\right)x^{-\alpha-1}(1-x)^{\theta-1} + \left(\alpha x+\theta(1-x)\right)(1-x)^{-\alpha-1}x^{\theta-1}\right).$$

Theorem 6. *Endow, as usual, $T_n^{\alpha,\theta}$ with the uniform probability measure on $\partial T_n^{\alpha,\theta}$. Then,*

$$\frac{1}{n^{\alpha}} T_n^{\alpha,\theta} \underset{n\to\infty}{\overset{(d)}{\to}} \mathcal{T}_{\alpha,\nu_{\alpha,\theta}}$$

for the rooted Gromov-Hausdorff-Prokhorov topology.

This result reinforces Proposition 2 of [33] which states the a.s. convergence of $T_n^{\alpha,\theta}$, in a certain finite dimensional sense, to a continuum fragmentation tree with parameters $\alpha, \nu_{\alpha,\theta}$. In view of Theorem 1, it suffices to check that hypothesis **(H)** holds, which in the present case states that for any $f : \mathcal{S}^{\downarrow} \to \mathbb{R}$ continuous and bounded with $|f(\mathbf{s})| \leq (1-s_1)$,

$$n^{\alpha} \sum_{k=\lceil n/2 \rceil}^{n-1} f\left(\frac{k}{n}, \frac{n-k}{n}, 0, \ldots\right) q_n\left((k,n-k,0,\ldots)\right) \to \int_{1/2}^{1} f(x,1-x,0,\ldots) f_{\alpha,\theta}(x)\mathrm{d}x\,.$$

To prove this, we use that $\int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, and rewrite $q_{\alpha,\theta}(n-1,k)$ as

$$q_{\alpha,\theta}(n-1,k) = \binom{n-1}{k} \frac{\alpha\,(n-1-k)+\theta k}{n-1} \frac{\Gamma(n-1+\theta-\alpha)}{\Gamma(1-\alpha)\Gamma(n-1+\theta)} \int_0^1 x^{k-\alpha-1}(1-x)^{n-k+\theta-2}\mathrm{d}x.$$

Then set for $x \in [0,1]$,

$$F(x) := f(x,1-x,0,\ldots)\mathbf{1}_{\{x>1/2\}} + f(1-x,x,0,\ldots)\mathbf{1}_{\{x\leq 1/2\}}$$

and note that $F(0) = 0$ and $|F(x)| \leq (1-x) \wedge x$, for every $x \in [0,1]$. We have,

$$\begin{aligned}
&\sum_{k=\lceil n/2 \rceil}^{n-1} f\left(\frac{k}{n}, \frac{n-k}{n}, 0, \ldots\right) q_n\left((k, n-k, 0, \ldots)\right) = \sum_{k=0}^{n-1} F\left(\frac{k}{n}\right) q_{\alpha,\theta}(n-1,k) \\
= \quad & \frac{\Gamma(n-1+\theta-\alpha)}{\Gamma(1-\alpha)\Gamma(n-1+\theta)} \int_0^1 \sum_{k=0}^{n-1} \binom{n-1}{k} \frac{\alpha(n-1-k)+\theta k}{n-1} F\left(\frac{k}{n}\right) x^{k-\alpha-1}(1-x)^{n-k+\theta-2} \mathrm{d}x \\
= \quad & \frac{\Gamma(n-1+\theta-\alpha)}{\Gamma(1-\alpha)\Gamma(n-1+\theta)} \int_0^1 \mathbb{E}\left[\left(\alpha\left(1-\frac{B_{n-1}^{(x)}}{n-1}\right)+\theta\frac{B_{n-1}^{(x)}}{n-1}\right) F\left(\frac{B_{n-1}^{(x)}}{n}\right)\right] x^{-\alpha-1}(1-x)^{\theta-1}\mathrm{d}x,
\end{aligned}$$

where $B_{n-1}^{(x)}$ denotes a binomial random variable with parameters $(n-1, x)$. We can assume that $B_{n-1}^{(x)}/n \to x$ a.s. on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and since F is continuous and bounded on $[0,1]$, we have

$$\mathbb{E}\left[\left(\alpha\left(1-\frac{B_{n-1}^{(x)}}{n-1}\right)+\theta\frac{B_{n-1}^{(x)}}{n-1}\right) F\left(\frac{B_{n-1}^{(x)}}{n}\right)\right] \to \left(\alpha\left(1-x\right)+\theta x\right) F(x), \qquad \text{for every } x \in [0,1].$$

Moreover,

$$\begin{aligned}
& \mathbb{E}\left[\left(\alpha\left(1-\frac{B_{n-1}^{(x)}}{n-1}\right)+\theta\frac{B_{n-1}^{(x)}}{n-1}\right) F\left(\frac{B_{n-1}^{(x)}}{n}\right)\right] \\
\leq \quad & \left((\alpha+\theta)\,\mathbb{E}\left[\frac{B_{n-1}^{(x)}}{n}\right]\right) \wedge \left(\alpha\mathbb{E}\left[1-\frac{B_{n-1}^{(x)}}{n-1}\right]+\theta\mathbb{E}\left[\frac{B_{n-1}^{(x)}}{n-1}\right]\right) \\
\leq \quad & \left((\alpha+\theta)\,x\right) \wedge \left(\alpha(1-x)+\theta x\right).
\end{aligned}$$

This is enough to conclude by dominated convergence that

$$\begin{aligned}
& \int_0^1 \mathbb{E}\left[\left(\alpha\left(1-\frac{B_{n-1}^{(x)}}{n-1}\right)+\theta\frac{B_{n-1}^{(x)}}{n-1}\right) F\left(\frac{B_{n-1}^{(x)}}{n}\right)\right] x^{-\alpha-1}(1-x)^{\theta-1}\mathrm{d}x \\
\underset{n\to\infty}{\longrightarrow} \quad & \int_0^1 \left(\alpha\left(1-x\right)+\theta x\right) F(x) x^{-\alpha-1}(1-x)^{\theta-1}\mathrm{d}x \\
= \quad & \Gamma(1-\alpha) \int_{1/2}^1 f(x, 1-x, \ldots) f_{\alpha,\theta}(x)\mathrm{d}x.
\end{aligned}$$

Last, Stirling's formula implies that

$$\frac{\Gamma(n-1+\theta-\alpha)}{\Gamma(n-1+\theta)} \underset{n\to\infty}{\sim} n^{-\alpha},$$

as wanted.

3 Preliminaries on self-similar fragmentations and trees

3.1 Partition-valued self-similar fragmentations

In this section, we recall the aspects of the theory of self-similar fragmentations that will be needed to prove Theorems 1 and 2. We refer the reader to [9] for more details.

3.1.1 Partitions of sets of integers

Let $B \subset \mathbb{N}$ be a possibly infinite, nonempty subset of the integers, and $\pi = \{\pi_1, \pi_2, \ldots\}$ be a partition of B. The (nonempty) sets $\pi_1, \pi_2, \ldots$ are called the *blocks* of π, we denote their number by $b(\pi)$. In order to lift the ambiguity in the labeling of the blocks, we will use, unless otherwise specified, the convention that $\pi_i, i \geq 1$ is defined inductively as follows: π_i is the block of π that contains the least integer of the set

$$B \setminus \bigcup_{j=1}^{i-1} \pi_j \,,$$

if the latter is not empty. For $i \in B$, we also let $\pi_{(i)}$ be the block of π that contains i.

We let $\mathcal{P}_B$ be the set of partitions of B. This forms a partially ordered set, where we let $\pi \preceq \pi'$ if the blocks of π' are all included in blocks of π (we also say that π' is finer than π). The minimal element is $\mathbb{O}_B = \{B\}$, and the maximal element is $\mathbb{I}_B = \{\{i\} : i \in B\}$.

If $B' \subseteq B$ is nonempty, the restriction of π to B', denoted by $\pi|_{B'}$ or $B' \cap \pi$ with a slight abuse of notations, is the element of $\mathcal{P}_{B'}$ whose blocks are the non-empty elements of $\{B' \cap \pi_1, B' \cap \pi_2, \ldots\}$.

If $B \subset \mathbb{N}$ is finite, with say n elements, then any partition $\pi \in \mathcal{P}_B$ with b blocks induces an element $\lambda(\pi) \in \mathcal{P}_n$ with b parts, given by the non-increasing rearrangement of the sequence $(\#\pi_1, \ldots, \#\pi_b)$.

A subset $B \subset \mathbb{N}$ is said to admit an asymptotic frequency if the limit

$$\lim_{n \to \infty} \frac{\#(B \cap [n])}{n}$$

exists. It is then denoted by $|B|$. It is a well-known fact, due to Kingman, that if π is a random partition of $\mathbb{N}$ with distribution invariant under the action of permutations (simply called exchangeable partition), then a.s. every block of π admits an asymptotic frequency. The law of π is then given by the paintbox construction of next section, for some probability measure ν. We let $|\pi|^{\downarrow} \in \mathcal{S}^{\downarrow}$ be the non-increasing rearrangement of the sequence $(|\pi_i|, i \geq 1)$. The exchangeable partition π is called *proper* if $\sum_{i=1}^{b(\pi)} |\pi_i| = 1$, which is equivalent to the fact that π has a.s. no singleton blocks.

3.1.2 Paintbox construction

Let ν be a dislocation measure, as defined around (2). We construct a σ-finite measure on $\mathcal{P}_{\mathbb{N}}$ by the so-called *paintbox construction*. Namely, for every $\mathbf{s} \in \mathcal{S}^{\downarrow}$ with $\sum_{i \geq 1} s_i = 1$, consider an i.i.d. sequence $(K_i, i \geq 1)$ such that

$$\mathbb{P}(K_1 = k) = s_k \,, \qquad k \geq 1 \,.$$

Then the partition π such that i, j are in the same block of π if and only if $K_i = K_j$ is exchangeable. We denote by $\rho_{\mathbf{s}}(\mathrm{d}\pi)$ its law. Note that $\rho_{\mathbf{s}}(\mathrm{d}\pi)$-a.s., it holds that $|\pi|^{\downarrow} = \mathbf{s}$, and π is a.s. proper under $\rho_{\mathbf{s}}$ if and only if $\mathbf{s}$ sums to 1. The measure

$$\kappa_\nu(\mathrm{d}\pi) := \int_{\mathcal{S}^{\downarrow}} \nu(\mathrm{d}\mathbf{s}) \rho_{\mathbf{s}}(\mathrm{d}\pi)$$

is a σ-finite measure on $\mathcal{P}_{\mathbb{N}}$, invariant under the action of permutations. From the integrability condition (2) on ν, it is easy to check that for $k \geq 2$, if

$$A_k = \{\pi \in \mathcal{P}_{\mathbb{N}} : \pi|_{[k]} \neq \{[k]\}\}$$

is the set of partitions whose trace on $[k]$ has at least two blocks, then

$$\kappa_\nu(A_k) = \int_{\mathcal{S}^{\downarrow}} \nu(\mathrm{d}\mathbf{s}) \Big(1 - \sum_{i \geq 1} s_i^k\Big) < \infty \,, \tag{8}$$

for every $k \geq 2$, since $1 - \sum_{i \geq 1} s_i^k \leq 1 - s_1^k \leq k(1 - s_1)$.

3.1.3 Exchangeable partitions of finite and infinite sets

In this section, we establish some elementary results concerning exchangeable partitions of $[n]$ or $\mathbb{N}$. The set of partitions with variable size, namely

$$\mathcal{P} = \mathcal{P}_{\mathbb{N}} \cup \bigcup_{n\geq 1} \mathcal{P}_{[n]} ,$$

is endowed with the distance

$$d_{\mathcal{P}}(\pi, \pi') = \exp(-\sup\{k \geq 1 : \pi|_{[k]} = \pi'|_{[k]}\}) .$$

In the sequel, convergence in distribution for partitions will be understood with respect to the separable and complete space $(\mathcal{P}, d_{\mathcal{P}})$. We will use the falling factorial notation

$$(x)_n = x(x-1)\dots(x-n+1) = \frac{\Gamma(x+1)}{\Gamma(x-n+1)}$$

for x a real number and $n \in \mathbb{N}$, $n < x+1$. When $x \in \mathbb{N}$, we extend the notation to all $n \in \mathbb{N}$, by setting $(x)_n = 0$ for $n \geq x+1$.

Lemma 1. *Let π be an exchangeable partition of $[n]$, and let $k \leq n$. Then for every $B \subset [k]$ with l elements such that $1 \in B$,*

$$\mathbb{P}\left([k] \cap \pi_{(1)} = B \,|\, \#\pi_{(1)}\right) = \frac{(\#\pi_{(1)} - 1)_{l-1}(n - \#\pi_{(1)})_{k-l}}{(n-1)_{k-1}} .$$

Proof. By exchangeability, the probability under consideration does depends on B only through its cardinality, and this equal to $\mathbb{P}\left(i_2, \dots, i_l \in \pi_{(1)}, j_1, \dots, j_{k-l} \notin \pi_{(1)} | \#\pi_{(1)}\right)$ for any pairwise disjoint $i_2, \dots, i_l, j_1, \dots, j_{k-l} \in \{2, 3, \dots, n\}$ (note that there are $\binom{n-1}{l-1}\binom{n-l}{k-l}$ such choices). Consequently,

$$\begin{aligned}
\mathbb{P}([k] \cap \pi_{(1)} = B \,|\, \#\pi_{(1)}) &= \frac{\mathbb{E}\left[\sum_{\substack{i_2,\dots,i_l \\ j_1,\dots,j_{k-l}}} \mathbf{1}_{\{i_2,\dots,i_l \in \pi_{(1)}\}} \mathbf{1}_{\{j_1,\dots,j_{k-l} \notin \pi_{(1)}\}} \,\Big|\, \#\pi_{(1)}\right]}{\binom{n-1}{l-1}\binom{n-l}{k-l}} \\
&= \frac{\binom{\#\pi_{(1)}-1}{l-1}\binom{n-\#\pi_{(1)}}{k-l}}{\binom{n-1}{l-1}\binom{n-l}{k-l}} ,
\end{aligned}$$

where the sum in the expectation is over indices considered above. This yields the result. $\square$

Lemma 2. *Let $(\pi^{(n)}, n \geq 1)$ be a sequence of random exchangeable partitions respectively in $\mathcal{P}_{[n]}$. We assume that $\pi^{(n)}$ converges in distribution to π. Then π is exchangeable and*

$$\frac{\#\pi^{(n)}_{(i)}}{n} \xrightarrow[n\to\infty]{(d)} |\pi_{(i)}| ,$$

the latter convergences holding jointly for $i \geq 1$, and jointly with the convergence $\pi^{(n)} \to \pi$.

Proof. The fact that π is invariant under the action of permutations of $\mathbb{N}$ with finite support is inherited from the exchangeability of $\pi^{(n)}$, and one concludes that π is exchangeable [5].

The random variables $(\#\pi^{(n)}_{(i)}/n, i \geq 1)$ take values in $[0, 1]$, so their joint distribution is tight, and up to extraction, we may assume that they converge in distribution to a random vector $(x_{(i)}, i \geq 1)$, jointly with the convergence $\pi^{(n)} \to \pi$. We want to show that a.s. $x_{(i)} = |\pi_{(i)}|$, which will characterize the limiting distribution.

For $k \geq l \geq 1$ fixed, by summing the formula of Lemma 1 over all $B \subset [k]$ containing i, with l elements, we get

$$\begin{aligned}\mathbb{P}\big(\#([k]\cap\pi^{(n)}_{(i)}) = l \,\big|\, \#\pi^{(n)}_{(i)}\big) &= \binom{k-1}{l-1}\frac{(\#\pi^{(n)}_{(i)}-1)_{l-1}(n-\#\pi^{(n)}_{(i)})_{k-l}}{(n-1)_{k-1}} \\ &\underset{n\to\infty}{\longrightarrow} \binom{k-1}{l-1} x_{(i)}^{l-1}(1-x_{(i)})^{k-l}\,,\end{aligned}$$

which entails that, conditionally on $x_{(i)}$, $\#([k]\cap\pi_{(i)})-1$ follows a binomial distribution with parameters $(k-1, x_{(i)})$. Therefore,

$$|\pi_{(i)}| = \lim_{k\to\infty}\frac{\#([k]\cap\pi_{(i)})}{k} = x_{(i)} \qquad \text{a.s.}\,,$$

by the law of large numbers. $\square$

Lemma 3. *Let $(\pi^{(i)}, 1 \leq i \leq r)$ be a sequence of random elements of $\mathcal{P}_{\mathbb{N}}$, which is exchangeable in the sense that $(\sigma\pi^{(i)}, 1 \leq i \leq r)$ has same distribution as $(\pi^{(i)}, 1 \leq i \leq r)$, for every permutation σ of $\mathbb{N}$. Then for every $k \geq 2$,*

$$\mathbb{P}\big(2,3,\ldots,k \in \pi^{(1)}_{(1)} \,\big|\, |\pi^{(i)}_{(j)}|, 1 \leq i \leq r, j \geq 1\big) = |\pi^{(1)}_{(1)}|^{k-1}\,.$$

Proof. Let $n \geq k$ and set $\pi^{(i,n)} = \pi^{(i)}|_{[n]}$, so that $(\pi^{(i,n)}, 1 \leq i \leq r)$ is a random sequence of $\mathcal{P}_{[n]}$ that is exchangeable. Then, by using the same argument as in the proof of Lemma 1, it holds that

$$\mathbb{P}\big(2,3,\ldots,k \in \pi^{(1,n)}_{(1)} \,\big|\, \#\pi^{(i,n)}_{(j)}, 1 \leq i \leq r, 1 \leq j \leq n\big) = \frac{(\#\pi^{(1,n)}_{(1)}-1)_{k-1}}{(n-1)_{k-1}}\,.$$

Using Lemma 2, and the fact that $(\pi^{(i,n)}, 1 \leq i \leq r)$ converges in distribution to $(\pi^{(i)}, 1 \leq i \leq r)$ as $n \to \infty$, it is then elementary to get the result by taking limits. $\square$

3.1.4 Poisson construction of homogeneous fragmentations

We now recall a useful construction of homogeneous fragmentations using Poisson point processes. We again fix a dislocation measure ν.

Consider a Poisson random measure $\mathcal{N}(\mathrm{d}t\mathrm{d}\pi\mathrm{d}i)$ on the set $\mathbb{R}_+ \times \mathcal{P}_{\mathbb{N}} \times \mathbb{N}$, with intensity measure $\mathrm{d}t \otimes \kappa_\nu(\mathrm{d}\pi) \otimes \#_{\mathbb{N}}(\mathrm{d}i)$, where $\#_{\mathbb{N}}$ is the counting measure on $\mathbb{N}$. We use a Poisson process notation $(\pi^0_t, i^0_t)_{t\geq 0}$ for the atoms of $\mathcal{N}$: for $t \geq 0$, if (t, π, i) is an atom of $\mathcal{N}$ then we let $(\pi^0_t, i^0_t) = (\pi, i)$, and if there is no atom of $\mathcal{N}$ of the form (t, π, i), then we set $\pi^0_t = \mathbb{O}_{\mathbb{N}}$ and $i^0_t = 0$ by convention. One constructs a process $(\Pi^0(t), t \geq 0)$ by letting $\Pi^0(0) = \mathbb{O}_{\mathbb{N}}$, and given that $\Pi^0(s), 0 \leq s < t$ has been defined, we let $\Pi^0(t)$ be the element of $\mathcal{P}_{\mathbb{N}}$ obtained from $\Pi^0(t-)$ by leaving its blocks unchanged, except the i^0_t-th block $\Pi^0_{i^0_t}(t-)$, which is intersected with π^0_t. Of course, this construction is only informal, since the times t of occurrence of an atom of $\mathcal{N}$ are everywhere dense in $\mathbb{R}_+$. However, using (8), it is possible to perform a similar construction for partitions restricted to $[k]$, and check that these constructions are consistent as k varies [9, Section 3.1.1]. The process $(\Pi^0(t), t \geq 0)$ is called a partition-valued homogeneous fragmentation with dislocation measure π.

Note in particular that the block $\Pi^0_{(1)}(t)$ that contains 1 at time t, is given by

$$\Pi^0_{(1)}(t) = \bigcap_{\substack{0<s\leq t\\ i^0_s=1}} (\pi^0_s)_{(1)}\,, \tag{9}$$

and that the restriction of $\mathcal{N}$ to $\mathbb{R}_+ \times \mathcal{P}_\mathbb{N} \times \{1\}$ is a Poisson measure with intensity $\mathrm{d}t \otimes \kappa_\nu(\mathrm{d}\pi)$.

For $k \geq 2$, let $D^0_k = \inf\{t \geq 0 : \Pi^0(t) \in A_k\}$ be the first time when the restriction of $\Pi^0(t)$ to $[k]$ has at least two blocks. By the previous construction, it is immediate to see that D^0_k has an exponential distribution with parameter $\kappa_\nu(A_k)$: it is the first time t such that $i^0_t = 1$ and $\pi^0_t \in A_k$. Moreover, by standard properties of Poisson random measures, conditionally on $D^0_k = s$, the random variables π^0_s and $(\pi^0_t, i^0_t)_{0\leq t<s}$ are independent, and the law of π^0_s equals $\kappa_\nu(\cdot \,|\, A_k) = \kappa_\nu(\cdot \cap A_k)/\kappa_\nu(A_k)$, while $(\pi^0_t, i^0_t)_{0\leq t<s}$ has same distribution as the initial process conditioned on $\{(\pi^0_t, i^0_t) \notin A_k \times \{1\}, 0 \leq t < s\} = \{D^0_k \geq s\}$, which has probability $e^{-s\kappa_\nu(A_k)}$. It is also equivalent to condition on $\{D^0_k > s\}$, since $\mathbb{P}(D^0_k = s) = 0$. The next statement sums up this discussion. By definition, we let $X(t \wedge s-) = X(t)\mathbf{1}_{\{t<s\}} + X(s-)\mathbf{1}_{\{t\geq s\}}$ for X càdlàg.

Lemma 4. *Let F, f be non-negative measurable functions. Then*

$$\begin{aligned}
&\mathbb{E}\Big[F(\Pi^0(t \wedge D^0_k-), t \geq 0) f(\pi^0_{D^0_k})\Big] \\
&\quad = \kappa_\nu(f|A_k) \int_0^\infty \kappa_\nu(A_k)\mathrm{d}s\, \mathbb{E}\big[F(\Pi^0(t \wedge s), t \geq 0)\, \mathbf{1}_{\{D^0_k > s\}}\big] .
\end{aligned}$$

Otherwise said, $\pi^0_{D^0_k}$ and $(\Pi^0(t \wedge D^0_k-), t \geq 0)$, are independent with respective laws $\kappa_\nu(\cdot \,|\, A_k)$, and the law of $(\Pi^0(t \wedge \mathbb{e}), t \geq 0)$ where $\mathbb{e}$ is an exponential random variable, independent of Π^0, and with parameter $\kappa_\nu(A_k)$.

3.1.5 Self-similar fragmentations

From a homogeneous fragmentation Π^0 constructed as above, one can associate a one-parameter family of $\mathcal{P}_\mathbb{N}$-valued processes by a time-changing method. Let $a \in \mathbb{R}$. For every $i \geq 1$ we let $(\tau^a_{(i)}(t), t \geq 0)$ be defined as the right-continuous inverse of the non-decreasing process

$$\int_0^t |\Pi^0_{(i)}(u)|^{-a} \mathrm{d}u \,, \qquad t \geq 0\,.$$

For $t \geq 0$, let $\Pi(t)$ be the random partition of $\mathbb{N}$ whose blocks are given by $\Pi^0_{(i)}(\tau^a_{(i)}(t)), i \geq 1$. One can check that this definition is consistent, namely, that for every $j \in \Pi^0_{(i)}(\tau^a_{(i)}(t))$, one has $\Pi^0_{(i)}(\tau^a_{(i)}(t)) = \Pi^0_{(j)}(\tau^a_{(j)}(t))$.

The process $(\Pi(t), t \geq 0)$ is called the self-similar fragmentation with index a and dislocation measure ν [9, Chapter 3.3]. We now assume that $a = -\gamma < 0$ is fixed once and for all. Let $D_k = \inf\{t \geq 0 : \Pi(t) \in A_k\}$.

Proposition 5. *Conditionally given $\sigma\{\Pi_{(i)}(t \wedge D_k) : t \geq 0, 1 \leq i \leq k\}$, and letting $\pi = \Pi(D_k)$, the random variable $(\Pi_i(t + D_k), t \geq 0)_{1\leq i\leq b([k]\cap\pi)}$ has same distribution as $(\pi_i \cap \Pi^{(i)}(|\pi_i|^a t), t \geq 0)_{1\leq i\leq b([k]\cap\pi)}$, where $(\Pi^{(i)}, i \geq 1)$ are i.i.d. copies of Π.*

Proof. For every $i \geq 1$ we let $L_i = \inf\{t \geq 0 : \Pi_{(i)}(t) \cap [k] \neq [k]\}$. Then $L = (L_i, i \geq 1)$ is a so-called *stopping line*, i.e. for every $i \geq 1$, L_i is a stopping time with respect to the natural filtration of $\Pi_{(i)}$, while $L_i = L_j$ for every $j \in \Pi_{(i)}(L_i)$. We let $\Pi(L)$ be the partition whose blocks are $\Pi_{(i)}(L_i), i \geq 1$ — by definition of a stopping line, two such blocks are either equal or disjoint. Note that $t + L = (t + L_i, i \geq 1)$ is also a stopping line, as well as $t \wedge L = (t \wedge L_i, i \geq 1)$.

From the so called *extended branching property* [9, Lemma 3.14], we obtain that conditionally given $\sigma\{\Pi(t \wedge L), t \geq 0\}$, the process $(\Pi(t + L), t \geq 0)$ has same distribution as

$$(\{\pi_i \cap \Pi^{(i)}(|\pi_i|^a t)), i \geq 1\}, t \geq 0)\,,$$

where $\pi = \Pi(L)$ and $(\Pi^{(i)}, i \geq 1)$ are i.i.d. copies of Π. The result is then a specialization of this, when looking only at the blocks of Π that contain $1, 2, \ldots, k$. $\square$

It will be of key importance to characterize the joint distribution of $D_k, (\Pi_{(i)}(D_k), 1 \leq i \leq k)$. This can be obtained as a consequence of Lemma 4. Recall the construction of Π from Π^0, let $\tau_{(i)} = \tau^a_{(i)}$, and define $\pi_t = \pi^0_{\tau_{(1)}(t)}$. The latter is equal to $\pi^0_{\tau_{(i)}(t)}$ for every $i \in [k]$ and $t \leq D_k$.

Proposition 6. *Let F, f be non-negative, measurable functions. Then*

$$\begin{aligned}
&\mathbb{E}\Big[F(|\Pi_{(1)}(t \wedge D_k-)|, t \geq 0) f(\pi_{D_k})\Big] \\
&\quad = \kappa_\nu(f|A_k) \int_0^\infty \mathrm{d}u \, \kappa_\nu(A_k) \mathbb{E}\left[|\Pi_{(1)}(u)|^{k-1+a} \mathbf{1}_{\{|\Pi_{(1)}(u)|>0\}} F(|\Pi_{(1)}(t \wedge u)|, t \geq 0)\right]
\end{aligned}$$

Proof. By definition, D_k (resp. D^0_k) is the first time when $[k] \cap \Pi(t) \neq [k]$ (resp. $[k] \cap \Pi^0(t) \neq [k]$). It follows that $D^0_k = \tau_{(1)}(D_k)$, and that the process

$$\Pi_{(1)}(t \wedge D_k-) = \Pi^0_{(1)}(\tau_{(1)}(t \wedge D_k-)) = \Pi^0_{(1)}(\tau_{(1)}(t) \wedge D^0_k-) \,, \qquad t \geq 0$$

is measurable with respect to $\sigma\{\Pi^0_{(1)}(t \wedge D^0_k-), t \geq 0\}$. Lemma 4 implies that

$$\begin{aligned}
&\mathbb{E}\Big[F(|\Pi_{(1)}(t \wedge D_k-)|, t \geq 0) f(\pi_{D_k})\Big] \\
&\quad = \mathbb{E}\Big[F(|\Pi^0_{(1)}(\tau_{(1)}(t) \wedge D^0_k-)|, t \geq 0) f(\pi^0_{D^0_k})\Big] \\
&\quad = \kappa_\nu(f|A_k) \int_0^\infty \mathrm{d}s \, \kappa_\nu(A_k) \mathbb{E}\Big[F\big(|\Pi^0_{(1)}(\tau_{(1)}(t) \wedge s)|, t \geq 0\big) \mathbf{1}_{\{D^0_k > s\}}\Big] \\
&\quad = \kappa_\nu(f|A_k) \mathbb{E}\Bigg[\int_0^\infty \mathrm{d}u \, \kappa_\nu(A_k) |\Pi_{(1)}(u)|^a F(|\Pi^0_{(1)}(\tau_{(1)}(t) \wedge \tau_{(1)}(u))|, t \geq 0) \mathbf{1}_{\{D_k > u\}}\Bigg] \\
&\quad = \kappa_\nu(f|A_k) \mathbb{E}\Bigg[\int_0^\infty \mathrm{d}u \, \kappa_\nu(A_k) |\Pi_{(1)}(u)|^a F(|\Pi_{(1)}(t \wedge u)|, t \geq 0) \mathbf{1}_{\{D_k > u\}}\Bigg] \,,
\end{aligned}$$

where in the third equality, we used successively Fubini's theorem and the change of variables $s = \tau_{(1)}(u)$, so that $\mathrm{d}s = |\Pi_{(1)}(u)|^a \mathrm{d}u$. We conclude by using the fact that

$$\mathbb{P}\Big(D_k > u \,\Big|\, |\Pi_{(1)}(t)|, 0 \leq t \leq u\Big) = |\Pi_{(1)}(u)|^{k-1} \,, \tag{10}$$

which can be argued as follows. Let $0 \leq t_1 < t_2 < \ldots < t_r = u$ be fixed times, then by applying Lemma 3 to the sequence $(\Pi(t_i), 1 \leq i \leq r)$, we obtain that

$$\mathbb{P}\Big(D_k > u \,\Big|\, |\Pi_{(1)}(t_i)|, 1 \leq i \leq r\Big) = |\Pi_{(1)}(u)|^{k-1} \,.$$

This yields (10) by a monotone class argument, using the fact that $\sigma\{|\Pi_{(1)}(t)|, 0 \leq t \leq u\}$ is generated by finite cylinder events. $\square$

The last important property of self-similar fragmentation is that the process $(|\Pi_{(1)}(t)|, t \geq 0)$ is a Markov process, which can be described as follows [9]. Let $(\xi_t, t \geq 0)$ be a subordinator with Laplace transform

$$\mathbb{E}\left[\exp(-r\xi_t)\right] = \exp\left(-t \int_0^\infty \Big(1 - \sum_{i \geq 1} s_i^{r+1}\Big) \nu(\mathrm{d}\mathbf{s})\right) .$$

Then $(|\Pi^0_{(1)}(t)|, t \geq 0)$ as same distribution as $(\exp(-\xi_t), t \geq 0)$, and consequently, the process $(|\Pi_{(1)}(t)|, t \geq 0)$ is a so-called self-similar Markov process:

Proposition 7 (Corollary 3.1 of [9])**.** *The process $(|\Pi_{(1)}(t)|, t \geq 0)$ has same distribution as $\exp(-\xi_{\tau(t)}, t \geq 0)$, where τ is the right-continuous inverse of the process $(\int_0^u \exp(a\xi_s)\mathrm{d}s, u \geq 0)$.*

3.2 Continuum fragmentation trees

This section is devoted to a more detailed description of the limiting self-similar fragmentation tree $\mathcal{T}_{\gamma,\nu}$ [24]. In particular, we will need a new decomposition result of reduced trees at the first branchpoint (Proposition 10).

3.2.1 Trees with edge-lengths and $\mathbb{R}$-trees

We saw in Section 1.3.1 how to turn a tree into a (finite) measured metric space. It is also easy to "turn discrete trees into $\mathbb{R}$-trees", viewing the edges as real segments of length 1.

More generally, a plane tree *with edge-length* is a pair $\theta = (\mathfrak{t}, (\ell_u, u \in \mathfrak{t}))$ where $\ell_u \geq 0$ for every $u \in \mathfrak{t}$, and a *tree with edge-lengths* is obtained by "forgetting the ordering" in a way that is adapted from the discussion of Section 1.1 in a straightforward way. Namely, the plane trees with edge-lengths $(\mathfrak{t}, (\ell_u, u \in \mathfrak{t}))$ and $(\mathfrak{t}', (\ell'_u, u \in \mathfrak{t}'))$ are equivalent if there exist permutations $\boldsymbol{\sigma} = (\sigma_u, u \in \mathfrak{t})$ such that $\boldsymbol{\sigma}\mathfrak{t} = \mathfrak{t}'$ and $\ell'_{\boldsymbol{\sigma}(u)} = \ell_u$, for every $u \in \mathfrak{t}$. We let $\boldsymbol{\Theta}$ be the set of trees with edge-lengths, i.e. of equivalence classes of plane trees with edge-lengths. There is a natural concatenation transformation, similar to $\langle \cdot \rangle$, for elements of $\boldsymbol{\Theta}$. Namely, if $\theta^{(i)} = (\mathfrak{t}^{(i)}, (\ell^{(i)}_u, u \in \mathfrak{t})), 1 \leq i \leq k$ is a sequence of plane trees with edge-lengths and $\ell \geq 0$, let

$$\langle \theta^{(i)}, 1 \leq i \leq k \rangle_\ell = (\mathfrak{t}, (\ell_u, u \in \mathfrak{t}))\,,$$

be defined by

$$\mathfrak{t} = \langle \mathfrak{t}^{(i)}, 1 \leq i \leq k \rangle\,,$$

and

$$\ell_\varnothing = \ell\,, \qquad \ell_{iu} = \ell^{(i)}_u\,, \quad 1 \leq i \leq k,\, u \in \mathfrak{t}^{(i)}\,.$$

If we replace each $\theta^{(i)}$ by another equivalent plane tree with edge-lengths, then the resulting concatenation is equivalent to the first one, so that this operation is well-defined for elements of $\boldsymbol{\Theta}$

Let $\theta \in \boldsymbol{\Theta}$, and consider a plane representative $(\mathfrak{t}, (\ell_u, u \in \mathfrak{t}))$. We construct an $\mathbb{R}$-tree $\mathcal{T}$ by imagining that the edge from $\mathrm{pr}(u)$ to u has length ℓ_u. Note that this intuitively involves a new edge with length $\ell_\varnothing$ pointing from the root ρ of the resulting $\mathbb{R}$-tree to $\varnothing$ (this is sometimes called *planting*). Formally, $\mathcal{T}$ is the isometry-equivalence class of a subset of $\mathbb{R}^{\mathfrak{t}}$ endowed with the l^1-norm $\|(x_u, u \in \mathfrak{t})\|_1 = \sum_{u\in\mathfrak{t}} |x_u|$, defined as the union of segments

$$\bigcup_{u\in\mathfrak{t}} \Big[\sum_{v\prec u} \ell_v e_v, \sum_{v\prec u} \ell_v e_v + \ell_u e_u\Big]\,,$$

where $(e_u, u \in \mathfrak{t})$ is the canonical basis of $\mathbb{R}^{\mathfrak{t}}$ and $v \prec u$ means that v is a strict ancestor of u in $\mathfrak{t}$. This $\mathbb{R}$-tree is naturally rooted at $0 \in \mathbb{R}^{\mathfrak{t}}$. Of course, its isometry class does not depend on the choice of the plane representative of θ, and can be written $\mathcal{T}(\theta)$ unambiguously. Note that there is a natural "embedding" mapping $\iota : \mathtt{t} \to \mathcal{T}(\theta)$ inherited from

$$\iota_0 : \mathfrak{t} \to \mathcal{T}\,, \qquad \iota_0(u) = \sum_{v\preceq u} \ell_v e_v\,, \tag{11}$$

and the latter is an isometry if θ is endowed with the (semi-)metric d_θ on its vertices, defined by

$$d_\theta(u, v) = \sum_{w\preceq u \,\mathrm{xor}\, w\preceq v} \ell_w\,,$$

where xor denotes "exclusive or".

Conversely, it is an elementary exercise to see that any rooted $\mathbb{R}$-tree $\mathcal{T}$ with a finite number of leaves can be written in the form $\mathcal{T} = \mathcal{T}(\theta)$ for some $\theta \in \boldsymbol{\Theta}$, which is in fact unique. In the sequel,

we will often identify the tree $\theta \in \mathbf{\Theta}$ with the $\mathbb{R}$-tree $\mathcal{T}(\theta)$. For instance, this justifies the notation $\langle \mathcal{T}^{(1)}, \ldots, \mathcal{T}^{(r)} \rangle_\ell$ for $\mathbb{R}$-trees $\mathcal{T}^{(1)}, \ldots, \mathcal{T}^{(r)}$ with finitely many leaves and for $\ell \geq 0$, which stands for the $\mathbb{R}$-tree in which the roots of $\mathcal{T}^{(1)}, \ldots, \mathcal{T}^{(r)}$ have been identified, and attached to a segment of length ℓ to a new root.

With a discrete tree t , we canonically associate the tree with edge-lengths θ in which all lengths are equal to 1, and the rooted $\mathbb{R}$-tree $\mathcal{T}(\mathsf{t}) = \mathcal{T}(\theta)$. In this case, $d_\theta = d_{\mathrm{gr}}$ is the graph distance. Using the isometry $\iota : \mathsf{t} \mapsto \mathcal{T}(\mathsf{t})$, we get the following statement, left as an exercise to the reader.

Proposition 8. *Viewing $\mathsf{t} \in \mathsf{T}$ as the element $(\mathsf{t}, d_{\mathrm{gr}}, \rho, \mu_{\partial \mathsf{t}})$ of $\mathscr{M}_{\mathrm{w}}$ as in Section 1.3.1, and endowing $\mathcal{T}(\mathsf{t})$ with the uniform probability distribution on $\mathcal{L}(\mathcal{T}(\mathsf{t}))$, it holds that*

$$d_{\mathrm{GHP}}(a\mathsf{t}, a\mathcal{T}(\mathsf{t})) \leq a\,, \qquad a > 0\,.$$

Due to this statement, in order to prove that the Markov branching tree T_n with law P^q_n converges after rescaling towards $\mathcal{T}_{\gamma,\nu}$, it suffices to show the same statement for the $\mathbb{R}$-tree $\mathcal{T}(T_n)$. We will often make the identification of T_n with $\mathcal{T}(T_n)$.

3.2.2 Partition-valued processes and $\mathbb{R}$-trees

Let $(\pi(t), t \geq 0)$ be a process with values in $C \subset \mathbb{N}$, finite or infinite, which is non-decreasing and indexed either by $t \in \mathbb{Z}_+$ or $t \in \mathbb{R}_+$, in which case we also assume that $\pi(\cdot)$ is right-continuous. We assume that there exists some $t_0 > 0$ such that $\pi(t_0) = \mathbb{I}_C$. Let $B \subseteq C$ be finite. If $B = \{i\}$, we let

$$D^\pi_{\{i\}} = \inf\{t \geq 0 : \{i\} \in \pi(t)\}$$

be the first time where i is isolated in a singleton block, and for $\#B \geq 2$, let

$$D^\pi_B = \inf\{t \geq 0 : B \cap \pi(t) \neq B\}\,.$$

We can build a tree with edge-lengths (and labeled leaves) $\theta(\pi(\cdot), B)$ by he following inductive procedure:

1. If $B = \{i\}$ we let $\theta(\pi(\cdot), B)$ be the tree $\bullet$ with length $D^\pi_{\{i\}}$
2. if $\#B \geq 2$, we let

$$\theta(\pi(\cdot), B) = \langle \theta(\pi(D^\pi_B + \cdot), B \cap \pi_i(D^\pi_B)), 1 \leq i \leq b \rangle_{D^\pi_B}\,,$$

   where b is the number of blocks of $\pi(D^\pi_B)$ that intersect B, and which are denoted by $\pi_1(D^\pi_B), \ldots, \pi_b(D^\pi_B)$.

Note that the previous labeling convention for blocks may not agree with our usual convention of labeling with respect to order of least element.

If $(\pi(t), t \in \mathbb{Z}_+)$ is indexed by non-negative integers, and satisfies $\pi(0) = \mathbb{O}_C$, there is a similar construction with trees rather than trees with edge-lengths. Namely, we let $\mathsf{t}_{\pi(\cdot)}$ be defined by

1. $\mathsf{t}_{\pi(\cdot)} = \bullet$ if $\#C = 1$, and
2. $\mathsf{t}_{\pi(\cdot)} = \langle \mathsf{t}_{\pi_i(1) \cap \pi(\cdot + 1)}, 1 \leq i \leq b \rangle$ otherwise, where b is the number of blocks of $\pi(1)$, denoted by $\pi_1(1), \ldots, \pi_b(1)$.

It is then easy to see that, with the notations of the previous section,

$$\mathcal{T}(\mathsf{t}_{\pi(\cdot)}) = \mathcal{T}(\theta(\pi(\cdot), C))\,, \tag{12}$$

And one can view $\theta(\pi(\cdot), B)$ as the subtree of $\mathsf{t}_{\pi(\cdot)}$ spanned by the root and the leaves with labels in B.

3.2.3 Continuum fragmentation trees

Let $(\Pi(t), t \geq 0)$ be the self-similar fragmentation process with index $-\gamma < 0$ and dislocation measure ν. The formation of dust property alluded to in Section 1.4 amounts to the fact that almost-surely, there exists some time $t_0 > 0$ such that $\Pi(t) = \mathbb{I}_{\mathbb{N}}$ for every $t \geq t_0$. Consequently, the construction of the previous paragraph applies with $C = \mathbb{N}$, and allows to construct a family of $\mathbb{R}$-trees

$$\mathcal{R}_B = \theta(\Pi(\cdot), B)$$

indexed by finite subsets $B \subset \mathbb{N}$. Recall that a tree $\theta \in \mathbf{\Theta}$ has been identified with $\mathcal{T}(\theta) \in \mathscr{T}_{\mathrm{w}}$. These $\mathbb{R}$-trees have finitely many leaves that are naturally indexed by elements of B. Moreover, they satisfy an obvious consistency property, meaning that taking the subtree spanned by the root and the leaves indexed by $B' \subset B$ yields an $\mathbb{R}$-tree with same law as $\mathcal{R}_{B'}$. This is the key to the definition of the fragmentation tree $\mathcal{T}_{\gamma,\nu}$.

Proposition 9 ([24])**.** *Conditionally given $\mathcal{T}_{\gamma,\nu} = (\mathcal{T}, d, \rho, \mu)$, let $L_1, L_2, \ldots$ be an i.i.d. sequence of leaves of $\mathcal{T}$ distributed according to μ. Then for every finite $B \subset \mathbb{N}$, the* reduced subtree

$$\mathcal{R}(\mathcal{T}_{\gamma,\nu}, B) = \bigcup_{i \in B} [[\rho, L_i]]$$

has same distribution as $\mathcal{R}_B$.

Moreover, the law of $\mathcal{T}_{\gamma,\nu}$ is the only one having this property, among distributions on $\mathscr{T}_{\mathrm{w}}$ that charge only the set of $\{(\mathcal{T}, d, \rho, \mu) \in \mathscr{T}_{\mathrm{w}} : \forall\, x \in \mathcal{T}, x \notin \mathcal{L}(\mathcal{T}) \implies \mu(\mathcal{T}_x) > 0\}$.

As an easy consequence, we have the following "converse construction" of fragmentations from $\mathcal{T}_{\gamma,\nu}$. With the notation of the proposition, for every $t \geq 0$, let $\Pi(t)$ be the partition of $\mathbb{N}$ such that i, j are in the same block of $\Pi(t)$ if and only if $d(\rho, L_i \wedge L_j) > t$. Then $(\Pi(t), t \geq 0)$ is a self-similar fragmentation process with dislocation measure ν and index $-\gamma$.

Also, note that the *reduced trees* $\mathcal{R}(\mathcal{T}_{\gamma,\nu}, B)$ rooted at ρ and endowed with the empirical measure

$$\mu_B = \frac{1}{\#B} \sum_{i \in B} \delta_{L_i},$$

converge in distribution as $\#B \to \infty$ in $\mathscr{T}_{\mathrm{w}}$ towards $(\mathcal{T}, d, \rho, \mu)$. In fact, the convergence holds a.s. if $B = [k]$ with $k \to \infty$: this is a simple exercise using the fact that $\{L_i, i \geq 1\}$ is a.s. dense in $\mathcal{L}(\mathcal{T})$ (by property 3. in the definition of $\mathcal{T}_{\gamma,\nu}$), and the weak convergence of $\mu_{[k]}$ to μ as $k \to \infty$.

The following statement gives a decomposition of the reduced tree $\mathcal{R}(\mathcal{T}, [k])$ at its first branchpoint above the root. Recall the notation $D_k = \inf\{t \geq 0 : \Pi(t) \in A_k\}$.

Proposition 10. *Let $k \geq 2$ and $\pi = \Pi(D_k), \pi' = \pi|_k, b = b(\pi')$. Then conditionally on $\{\pi, D_k\}$, the reduced tree $\mathcal{R}(\mathcal{T}_{\gamma,\nu}, [k])$ has same distribution as*

$$\mathcal{T}\Big(\big\langle |\pi_i|^\gamma \mathcal{R}(\mathcal{T}^{(i)}, \pi'_i), 1 \leq i \leq b \big\rangle_{D_k}\Big),$$

where the $\mathcal{T}^{(i)}$ are i.i.d. with same distribution as $\mathcal{T}_{\gamma,\nu}$, independent of $\sigma\{\pi, D_k\}$.

Moreover, for every $i \in \mathbb{N}$, the tree $\mathcal{R}(\mathcal{T}_{\gamma,\nu}, \{i\})$ has same distribution as the $\mathbb{R}$-tree associated with the tree $(\varnothing, D_1) \in \mathbf{\Theta}$, i.e. a real segment with length $D_1 = \inf\{t \geq 0 : \{1\} \in \Pi(t)\}$.

Proof. The second statement is just a matter of definitions, so we only need to prove the first one. By Proposition 5, the process $\Pi(D_k + \cdot)$, in restriction to the blocks containing at least one element in $[k]$, has same distribution as the partitions-valued process whose blocks are those of

$\pi_i \cap \Pi^{(i)}(|\pi_i|^{-\gamma}\cdot), 1 \leq i \leq b$, for i.i.d. copies $\Pi^{(i)}, i \geq 1$ of Π, independent on π, D_k. Therefore, one gets from the definition of $\mathcal{R}_B$ that

$$\mathcal{R}_{[k]} \stackrel{(d)}{=} \mathcal{T}\langle \theta(\Pi^{(i)}(|\pi_i|^{-\gamma}\cdot), \pi'_i), 1 \leq i \leq b\rangle_{D_k},$$

from which the result follows immediately. □

Note that Proposition 6 gives the joint distribution of $D_k, |\pi_i|, 1 \leq i \leq b, \pi'$ as a special case, while Proposition 7 characterizes the law of D_1, since it is the first time where the process $(|\Pi_{(1)}(t)|, t \geq 0)$ attains 0. This, together with the previous proposition, allows to characterize entirely the laws of the reduced trees of $\mathcal{T}_{\gamma,\nu}$, hence the law of $\mathcal{T}_{\gamma,\nu}$ itself.

3.2.4 Markov branching trees as discrete fragmentation trees

Recall the informal description of Markov branching trees P_n^q in the introduction, relying on collections of balls in urns. Rather than collections of indistinguishable balls that split randomly, it is convenient to consider instead a collection of balls that are distinguished by a random, exchangeable labeling. This is achieved by replacing partitions of integers by partitions of sets. We start with a preliminary lemma.

Lemma 5. *Let $n \geq 1$ be fixed, as well as a partition $\lambda \in \mathcal{P}_n$ with $p = p(\lambda)$ parts.*
(i) *There are*

$$C_\lambda = \frac{n!}{\prod_{i=1}^p \lambda_i! \prod_{j=1}^n m_j(\lambda)!}$$

partitions $\pi \in \mathcal{P}_{[n]}$ such that $\lambda(\pi) = \lambda$.
(ii) *If $1 \leq k \leq n$ and $\pi' \in \mathcal{P}_{[k]}$ has b blocks, then for every $i_1, \ldots, i_b \in \{1, 2, \ldots, p\}$ pairwise distinct, there are*

$$C_\lambda^{\pi'}(i_1, \ldots, i_b) = C_\lambda \frac{1}{(n)_k} \prod_{j=1}^b (\lambda_{i_j})_{\#\pi'_j}$$

partitions $\pi \in \mathcal{P}_{[n]}$ such that $\lambda(\pi) = \lambda$, $\pi|_{[k]} = \pi'$ and $\#\pi_j = \lambda_{i_j}, 1 \leq j \leq b$.

Proof. Let p be the number of parts of λ. Then there are $p!/\prod_{j=1}^n m_j(\lambda)!$ sequences $(c_1, \ldots, c_p)$ whose non-increasing rearrangement is λ. With any such sequence, we can associate

$$\frac{n!}{\prod_{i=1}^p c_i!} = \frac{n!}{\prod_{i=1}^p \lambda_i!}$$

sequences of the form $(B_1, \ldots, B_p)$ such that $\{B_1, \ldots, B_p\}$ is a partition of $[n]$ (beware that the labeling of the blocks B_i will differ in general from labeling convention described above for the blocks of a partition), with $\#B_i = c_i, 1 \leq i \leq p$. Finally, exactly $p!$ sequences of the form $(B_1, \ldots, B_p)$ induce the same partition $\{B_1, \ldots, B_p\}$. Putting things together easily yield the formula for C_λ.

For the second formula, if $\lambda \in \mathcal{P}_n, \pi' \in \mathcal{P}_{[k]}$ and $i_1, \ldots, i_b$ are given with $b = b(\pi')$, then any partition $\pi \in \mathcal{P}_{[n]}$ with $\lambda(\pi) = \lambda$ and $\pi|_{[k]} = \pi'$ must have $\pi_i|_{[k]} = (\pi|_{[k]})_i = \pi'_i$, for $1 \leq i \leq b$, the first equality coming from our choice of the labeling of blocks of partitions. The restriction of π to $[k]$ is thus entirely determined, while the restriction of π to $[n] \setminus [k]$ is a partition of the latter set whose block-sizes are given by the sequence $\lambda_i, i \notin \{i_1, \ldots, i_b\}, \lambda_{i_j} - \#\pi'_j, 1 \leq j \leq b$. A simple adaptation of point (i) shows that there are

$$\frac{(n-k)!}{\prod_{i \notin \{i_1, \ldots, i_b\}} \lambda_i! \prod_{j=1}^b (\lambda_{i_j} - \#\pi'_j)!} = C_\lambda^{\pi'}(i_1, \ldots, i_b)$$

such partitions. □

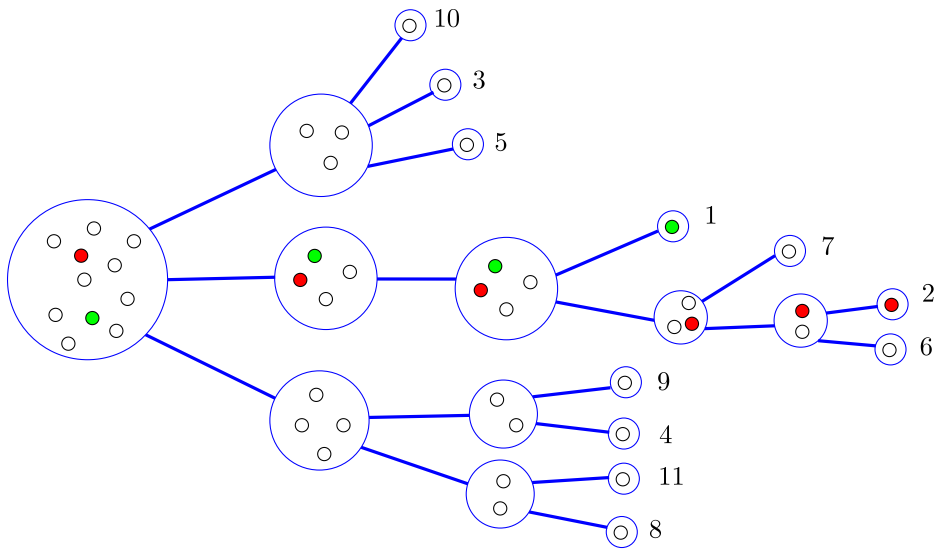

Figure 2: An sample tree $T_{[n]}$ for $n = 11$, with the labeled leaves. The process $\Pi^{(11)}$ can be easily deduced: for instance, $\Pi^{(11)}(1) = \{\{1, 2, 6, 7\}, \{3, 5, 10\}, \{4, 8, 9, 11\}\}$. As opposed to Figure 1, leaves are all connected to vertices with at least 2 children, because of the requirement $q_1(\emptyset) = 1$.

Going back to Markov branching trees, let $B \subset \mathbb{N}$ have $n \geq 2$ elements. Let $q = (q_n, n \geq 1)$ satisfy (1), and also assume that $q_1(\emptyset) = 1$. For every $\pi \in \mathcal{P}_B$, set

$$p_B(\pi) = \frac{q_n(\lambda(\pi))}{C_{\lambda(\pi)}}, \tag{13}$$

where C_λ is the constant appearing in Lemma 5. Given the partition of n that it induces (which has distribution q_n), a p_B-distributed partition is thus uniform among possible choices of partitions of B. In particular, a random partition with distribution p_B is exchangeable, i.e. its law is invariant under the action of permutations of B. By convention, the law p_B, if $B = \{i\}$ is a singleton, is the Dirac mass at the partition $\{\{i\}\}$.

For every $\pi \in \mathcal{P}_B$ with blocks $\pi_1, \pi_2, \dots, \pi_k$ say, consider random partitions $\widetilde{\pi}^i, 1 \leq i \leq k$ of $\pi_1, \dots, \pi_k$ respectively, chosen independently with respective distributions $p_{\pi_1}, \dots, p_{\pi_k}$. We let $Q(\pi, \cdot)$ be the law of the partition $\bigcup_{1 \leq i \leq k} \widetilde{\pi}^i \in \mathcal{P}_B$. Then Q is the transition kernel of a Markov chain on $\mathcal{P}_B$ (for any finite $B \subset \mathbb{N}$), that ends at the state $\mathbb{I}_B$. It is easily seen that this Markov chain is exchangeable as a process. Moreover, the chain started from the state $\{B\}$, with $\#B = n$ has same distribution as the image of the chain started from $[n]$ under the action of any bijection $[n] \to B$.

For finite $C \subset \mathbb{N}$, we let $(\Pi^C(r), r \geq 0)$ be the chain with transition matrix Q and started from $\Pi^C(0) = \mathbb{O}_C$. Plainly, Π^C is non-decreasing and attains $\mathbb{I}_C$ in finite time a.s., so the construction of Section 3.2.2 applies and yields a family $\theta(\Pi^C(\cdot), B) \in \boldsymbol{\Theta}, B \subseteq C$, as well as a tree $T_C := \mathsf{t}_{\Pi^C(\cdot)}$. By construction, given that $\Pi^C(1)$ has blocks $\pi_1, \dots, \pi_b$, the trees $\mathsf{t}_{\pi_i \cap \Pi^C(\cdot+1)}, 1 \leq i \leq b$ are independent with same distribution as $T_{\pi_i}, 1 \leq i \leq b$ respectively. Since the law of the non-increasing rearrangement of $\#\pi_i, 1 \leq i \leq b$ is $q_{\#C}$, we readily obtain the following statement[2].

Lemma 6. *The tree T_C has law $\mathsf{P}^q_{\#C}$.*

[2]There is one subtlety in this statement, which is in the case $C = \{i\}$ for some $i \in \mathbb{N}$. Indeed, by construction we have $T_C = \bullet$ a.s., and this is the only place where we have to require that $q_1(\emptyset) = 1$.

In fact, the leaves of the tree T_C are naturally labeled by elements of C. We will use it in the sequel, without further formalizing the notion of trees with labeled leaves.

We will also use the shorthand notation T_C^B for the *reduced tree* $\theta(\Pi^C(\cdot), B)$. Using the above description, and applying the Markov property for Π^C at time $D_B^{\Pi^C}$ and the particular form of the Markov kernel Q, we immediately obtain the following, in the particular case $B = [k], C = [n]$.

Proposition 11. *Let $2 \le k \le n$. Then, conditionally on $D_{[k]}^{\Pi^{[n]}} = \ell$ and $\Pi^{[n]}(D_{[k]}^{\Pi^{[n]}}) = \pi$, with $\pi|_{[k]} = \pi'$, it holds that $T_{[n]}^{[k]}$ has same distribution as*

$$\langle \theta^{(i)}, 1 \le i \le b(\pi') \rangle_\ell \,,$$

where $\theta^{(i)}, 1 \le i \le b$ are independent with respective laws that of $T_{\pi_i}^{\pi_i'}, 1 \le i \le b(\pi')$.

Proof. The only subtle point is that $[k] \cap \pi_i = \pi_i', 1 \le i \le b$, since the labeling of the blocks of π, π' could differ. But since these partitions are respectively of $[n]$ and $[k]$, this cannot be the case. $\square$

4 Proofs of Theorems 1 and 2

Let $q = (q_n, n \ge 1)$ be a sequence of laws on $\mathcal{P}_n$ respectively, that satisfies (1) and **(H)**, for some fragmentation pair $(-\gamma, \nu)$ and some slowly varying function ℓ. In order to lighten notations, we let $a_n = n^\gamma \ell(n)$.

As we noticed in the introduction, it is easy to pass from the situation where $q_1(\emptyset) = 1$ to the general situation, by adding independent linear strings with Geometric($q_1(\emptyset)$)-distributed lengths to the n leaves of T_n. Since geometric distributions have exponential tails, the longest of these n strings will have a length at most $C \log n$ with probability going to 1 as $n \to \infty$, for some $C > 0$. If we let T_n^1 be the tree for which $q_1(\emptyset) > 0$ and T_n^2 the one for which $q_1(\emptyset) = 1$, coupled in the way depicted above, we easily get, for any $\gamma > 0$,

$$\mathbb{P}(d_{\mathrm{GHP}}(a_n^{-1} T_n^1, a_n^{-1} T_n^2) \le C a_n^{-1} \log n) \underset{n\to\infty}{\longrightarrow} 1\,.$$

Thus, we can deduce the convergence in distribution of $a_n^{-1} T_n^2$ to $\mathcal{T}_{\gamma,\nu}$ from that of $a_n^{-1} T_n^1$. Therefore, from now on and until the end of the present section, we make the following hypothesis, which will allow us to apply Lemma 6:

(H') The sequence $(q_n, n \ge 1)$ satisfies **(H)** and $q_1(\emptyset) = 1$.

4.1 Preliminary convergence lemmas

We now establish a couple of intermediate convergence results for the discrete model. Recall that the sequence of distributions $q_n, n \ge 2$ on $\mathcal{P}_n$ respectively induce distributions p_B on $\mathcal{P}_B$ for finite B by formula (13). By convention we set $p_n = p_{[n]}$.

Lemma 7. *Let $k \ge 2$ and let π' be an element in $\mathcal{P}_{[k]}$ with b blocks, $b \ge 2$. Let $g : (0, \infty)^b \to \mathbb{R}$ be a continuous function with compact support. Then, under assumption* **(H')***,*

$$a_n p_n \left(g\Big(\frac{\#\pi_1}{n}, \dots, \frac{\#\pi_b}{n}\Big) \mathbf{1}_{\{\pi|_{[k]} = \pi'\}} \right) \underset{n\to\infty}{\longrightarrow} \int_{\mathcal{P}_\mathbb{N}} \kappa_\nu(\mathrm{d}\pi) g(|\pi_1|, \dots, |\pi_b|) \mathbf{1}_{\{\pi|_{[k]} = \pi'\}} \,,$$

where κ_ν is the paintbox construction associated with ν. Note that on the event $\{\pi|_{[k]} = \pi'\}$, the quantities $\#\pi_i/n$ and $|\pi_i|$ for $1 \le i \le b$ that appear above are a.e. non-zero, respectively under p_n and κ_ν.

Proof. For simplicity, we let

$$A_n = p_n\Big(g\Big(\frac{\#\pi_1}{n},\dots,\frac{\#\pi_b}{n}\Big)\mathbf{1}_{\{\pi|_{[k]}=\pi'\}}\Big)\,.$$

By the definition of q_n and Lemma 5,

$$\begin{aligned} A_n &= \sum_{\lambda\in\mathcal{P}_n} q_n(\lambda) \sum_{\substack{i_1,\dots,i_b\geq 1\\ \text{pairwise distinct}}} g\Big(\frac{\lambda_{i_1}}{n},\dots,\frac{\lambda_{i_b}}{n}\Big)\frac{C^{\pi'}_\lambda(i_1,\dots,i_b)}{C_\lambda} \\ &= \sum_{\lambda\in\mathcal{P}_n} q_n(\lambda)\frac{1}{(n)_k} \sum_{\substack{i_1,\dots,i_b\geq 1\\ \text{pairwise distinct}}} g\Big(\frac{\lambda_{i_1}}{n},\dots,\frac{\lambda_{i_b}}{n}\Big)\prod_{j=1}^b(\lambda_{i_j})_{\#\pi'_j}\,. \end{aligned}$$

Now, the function

$$h(\mathbf{s}) = \sum_{\substack{i_1,\dots,i_b\geq 1\\ \text{pairwise distinct}}} g(s_{i_1},\dots,s_{i_b})\prod_{j=1}^b s_{i_j}^{\#\pi'_j}\,,\qquad \mathbf{s}\in\mathcal{S}^\downarrow$$

is continuous and bounded, because g is compactly supported in $(0,\infty)^b$, so that the sum is really a finite sum. Moreover,

$$h(\mathbf{s}) \leq K \sum_{\substack{0\leq k_1,k_2,\dots<k\\ k_1+k_2+\dots=k}} \frac{k!}{\prod_{j\geq 1}k_j!}\prod_{j\geq 1}s_j^{k_j} = K\Big(1-\sum_{j\geq 1}s_j^k\Big)\leq kK(1-s_1)\,, \tag{14}$$

where K is an upper-bound of $|g|$, and for every $\lambda\in\mathcal{P}_n$, it is easily checked that for large n, if $\varepsilon>0$ is such that $g(x_1,\dots,x_b)=0$ as soon as $\min_{1\leq i\leq b}x_i\leq\varepsilon$,

$$\Big(1-\frac{k}{\varepsilon n}\Big)^k h(\lambda/n)\leq \frac{1}{(n)_k}\sum_{\substack{i_1,\dots,i_b\geq 1\\ \text{pairwise distinct}}} g\Big(\frac{\lambda_{i_1}}{n},\dots,\frac{\lambda_{i_b}}{n}\Big)\prod_{j=1}^b(\lambda_{i_j})_{\#\pi'_j}\leq\Big(\frac{n}{n-k}\Big)^k h(\lambda/n)\,.$$

Letting $n\to\infty$ and applying **(H')**, which is validated by (14),

$$\lim_{n\to\infty} a_n A_n = \int_{\mathcal{S}^\downarrow}\nu(\mathrm{d}\mathbf{s})h(\mathbf{s}) = \int_{\mathcal{P}_\mathbb{N}}\kappa_\nu(\mathrm{d}\pi)g(|\pi_1|,\dots,|\pi_b|)\mathbf{1}_{\{\pi|_{[k]}=\pi'\}}\,,$$

the latter equality being a simple consequence of the paintbox construction of Section 3.1.2. □

Now, we associate with $(q_n, n\geq 1)$ a family of process $(\Pi^B(r), r\geq 0)$ with values in $\mathcal{P}_B$, as in Section 3.2.4. We let $\Pi^n=\Pi^{[n]}$ for simplicity, and set

$$D^n_k = D^{\Pi^n}_{[k]} = \inf\{r\geq 0: [k]\cap\Pi^n(r)\neq\{[k]\}\}$$

for $2\leq k\leq n$, and $D^n_1 = D^{\Pi^n}_{\{1\}} = \inf\{r\geq 0:\{1\}\in\Pi^n(r)\}$. Also, for $r\geq 0$ we let

$$X_n(r) = \#\Pi^n_{(1)}(r)\,.$$

Lemma 8. *Let $n,k\in\mathbb{N}$ be fixed, with $n\geq k\geq 2$, and let $\pi'\in\mathcal{P}_{[k]}$ have $b\geq 2$ blocks. Let F, f be measurable non-negative functions. Then*

$$\begin{aligned} &\mathbb{E}\Big[F\Big(X_n(\,\cdot\wedge(D^n_{[k]}-1))\Big)f(\#\Pi^n_i(D^n_{[k]}),1\leq i\leq b)\mathbf{1}_{\{[k]\cap\Pi^n(D^n_{[k]})=\pi'\}}\Big] \\ &= \sum_{r'>0}\mathbb{E}\Big[\frac{(X_n(r'-1)-1)_{k-1}}{(n-1)_{k-1}}F\big(X_n(\cdot\wedge(r'-1))\big)\,p_{X_n(r'-1)}(f(\#\pi_i,1\leq i\leq b)\mathbf{1}_{\{\pi|_{[k]}=\pi'\}})\Big] \end{aligned}$$

Proof. We first consider an expression of a more general form. For non-negative functions G, g, we have, using the Markov property at time $r'-1$ in the second step,

$$
\begin{aligned}
&\mathbb{E}\Big[G\big(\Pi^n(\cdot \wedge (D^n_{[k]}-1))\big)g\big(\Pi^n_{(1)}(D^n_{[k]}-1)\cap \Pi^n(D^n_{[k]})\big)\Big] \\
&= \sum_{r'>0}\mathbb{E}\Big[G\big(\Pi^n(\cdot\wedge(r'-1))\big)\mathbf{1}_{\{[k]\subset\Pi^n_{(1)}(r'-1)\}}g\big(\Pi^n_{(1)}(r'-1)\cap\Pi^n(r')\big)\mathbf{1}_{\{[k]\cap\Pi^n(r')\neq\{[k]\}\}}\Big] \\
&= \sum_{r'>0}\mathbb{E}\Big[G\big(\Pi^n(\cdot\wedge(r'-1))\big)\mathbf{1}_{\{[k]\subset\Pi^n_{(1)}(r'-1)\}}p_{\Pi^n_{(1)}(r'-1)}\big(g(\pi)\mathbf{1}_{\{[k]\cap\pi\neq\{[k]\}\}}\big)\Big]
\end{aligned}
$$

Specializing this formula to G depending only on X_n and $g(\pi)=f(\#\pi_1,\ldots,\#\pi_b)\mathbf{1}_{\{\pi|_{[k]}=\pi'\}}$, and using obvious exchangeability properties, we obtain

$$
\begin{aligned}
&\mathbb{E}\Big[F\big(X_n(\cdot\wedge(D^n_{[k]}-1))\big)f\big(\#\Pi^n_i(D^n_{[k]}),1\leq i\leq b\big)\mathbf{1}_{\{[k]\cap\Pi^n(D^n_{[k]})=\pi'\}}\Big] \\
&= \sum_{r'>0}\mathbb{E}\Big[F\big(X_n(\cdot\wedge(r'-1))\big)p_{X_n(r'-1)}(f(\#\pi_i,1\leq i\leq b)\mathbf{1}_{\{\pi|_{[k]}=\pi'\}})\mathbf{1}_{\{[k]\subset\Pi^n_{(1)}(r'-1)\}}\Big]
\end{aligned}
$$

All the terms in the expectation depend on $(X_n(r),0\leq r\leq r'-1)$, except the last one which is a function of $\Pi^n_{(1)}(r'-1)$. But by Lemma 1 (in fact, the variant used in the proof of Lemma 3),

$$
\mathbb{P}\big([k]\subset\Pi^n_{(1)}(r'-1)\,\big|\,(X_n(r),0\leq r\leq r'-1)\big)=\frac{(X_n(r'-1)-1)_{k-1}}{(n-1)_{k-1}}\,,
$$

giving the result. $\square$

In the sequel, $\Pi(\cdot)$ will denote a continuous-time self-similar fragmentation with characteristic pair $(-\gamma,\nu)$, and $D_k, k\geq 1$ will be defined as in Section 3.1.5.

Lemma 9. *Under assumption* **(H')***, it holds that*

$$
\Big(\frac{X_n(\lfloor a_n t\rfloor)}{n},t\geq 0\Big)\xrightarrow[n\to\infty]{(d)}(|\Pi_{(1)}(t)|,t\geq 0)\,,
$$

in distribution for the Skorokhod topology, jointly with the convergence

$$
\frac{1}{a_n}D^n_1\xrightarrow[n\to\infty]{(d)}D_1\,.
$$

Proof. For $n>k\geq 1$, let $p_{n,k}=\mathbb{P}(X_n(1)=k)$. Note that the process X_n is a non-increasing Markov chain started from n, with probability transitions $p_{i,j}, 1\leq j\leq i$. Then by a simple exchangeability argument,

$$
p_{n,k}=\sum_{\pi\in\mathcal{P}_{[n]}}p_n(\pi)m_k(\pi)\frac{k}{n}=\sum_{\lambda\in\mathcal{P}_n}q_n(\lambda)m_k(\lambda)\frac{k}{n},\quad 1\leq k\leq n,
$$

where $m_k(\pi)=m_k(\lambda(\pi))$ is the number of blocks of π with size k. Consider the associated generating function for $x\geq 0$

$$
F_n(x)=\sum_{k=1}^n\Big(\frac{k}{n}\Big)^x p_{n,k}=\sum_{\lambda\in\mathcal{P}_n}q_n(\lambda)\sum_{k=1}^n m_k(\lambda)\Big(\frac{k}{n}\Big)^{x+1}=\sum_{\lambda\in\mathcal{P}_n}q_n(\lambda)\sum_{i\geq 1}\Big(\frac{\lambda_i}{n}\Big)^{x+1}.
$$

Hence, $1-F_n(x)=\overline{q}_n(f)$, where $f(\mathbf{s})=1-\sum_{i\geq 1}s_i^{x+1}$, and by **(H')**,

$$
a_n(1-F_n(x))\underset{n\to\infty}{\longrightarrow}\int_{\mathcal{S}^\downarrow}\Big(1-\sum_{i\geq 1}s_i^{x+1}\Big)\nu(\mathrm{d}\mathbf{s})\,.
$$

This is exactly what we need to use [25, Theorem 1], stating that $(n^{-1}X_n(\lfloor a_n t\rfloor), t \geq 0)$ converges in distribution to the self-similar Markov process $\exp(-\xi_{\tau(\cdot)})$, as defined around Proposition 7. Moreover, this convergence holds jointly with the convergence of absorption times at 1, so $a_n^{-1}D_1^n$ converges to the absorption time at 0 of $\exp(-\xi_{\tau(\cdot)})$. By Proposition 7, the process $\exp(-\xi_{\tau(\cdot)})$ has same distribution as $(|\Pi_{(1)}(t)|, t \geq 0)$, which reaches 0 for the first time at time D_1. Hence the result. $\square$

Finally, the combination of the last two lemmas gives the last of our preliminary ingredients.

Lemma 10. *The following joint convergence in distribution holds:*

$$\left(\frac{D_k^n}{a_n}, [k] \cap \Pi^n(D_k^n), \left(\frac{\#\Pi_{(i)}^n(D_k^n)}{n}, i \in [k]\right)\right) \underset{n\to\infty}{\overset{(d)}{\longrightarrow}} \left(D_k, [k] \cap \Pi(D_k), (|\Pi_{(i)}(D_k)|, i \in [k])\right).$$

Proof. Let $\pi' \in \mathcal{P}_k$ have $b \geq 2$ blocks, and $f, g : (0,\infty) \to \mathbb{R}, h : (0,\infty)^b \to \mathbb{R}$ be continuous functions with compact support. Then by Lemma 8,

$$\begin{aligned}
&\mathbb{E}\Big[f\Big(\frac{D_k^n}{a_n}\Big)g\Big(\frac{X_n(D_k^n-1)}{n}\Big)h\Big(\frac{\#\Pi_i^n(D_k^n)}{X_n(D_k^n-1)}, 1 \leq i \leq b\Big)\mathbf{1}_{\{[k]\cap\Pi^n(D_k^n)=\pi'\}}\Big] \\
&= \sum_{r'>0} f\Big(\frac{r'}{a_n}\Big)\mathbb{E}\Big[\frac{(X_n(r'-1)-1)_{k-1}}{(n-1)_{k-1}} g\Big(\frac{X_n(r'-1)}{n}\Big)p_{X_n(r'-1)}\Big(h\Big(\frac{\#\pi_i}{X_n(r'-1)}, 1 \leq i \leq b\Big)\mathbf{1}_{\{\pi|_{[k]}=\pi'\}}\Big)\Big] \\
&= \frac{1}{a_n}\sum_{r'>0} f\Big(\frac{r'}{a_n}\Big)\mathbb{E}\Big[\Phi(n, X_n(r'-1))g\Big(\frac{X_n(r'-1)}{n}\Big)\Psi(X_n(r'-1))\Big] \\
&= \int_{1/a_n}^{\infty} f\Big(\frac{\lfloor a_n u\rfloor}{a_n}\Big)\mathrm{d}u\,\mathbb{E}\Big[\Phi(n, X_n(\lfloor a_n u\rfloor-1))g\Big(\frac{X_n(\lfloor a_n u\rfloor-1)}{n}\Big)\Psi(X_n(\lfloor a_n u\rfloor-1))\Big]
\end{aligned}$$

where

$$\Phi(n,x) = \frac{(x-1)_{k-1}}{(n-1)_{k-1}}\frac{a_n}{a_x} \underset{(n,x/n)\to(\infty,c)}{\longrightarrow} c^{k-1-\gamma},$$

and

$$\Psi(m) = a_m p_m\Big(h\Big(\frac{\#\pi_i}{m}, 1 \leq i \leq b\Big)\mathbf{1}_{\{\pi|_{[k]}=\pi'\}}\Big).$$

Note that the Potter's bounds for regularly varying functions [11, Th.1.5.6.] imply that $\Phi(n,x) \leq C\left(\frac{x}{n}\right)^{k-1-\gamma-1}$ for all $n \geq x \geq A$ for some finite positive constants C, A. In particular there exists some n_0 such that $\sup_{n\geq n_0, 0<x\leq n}\Phi(n,x)g(x/n) < \infty$ (since g is null in a neighborhood of 0). The joint use of Lemmas 7 and 9 entails by dominated convergence that the expectation term in the integral converges to (note that the quantities $|\pi_i|, 1 \leq i \leq b$ are all a.e. positive on $\{\pi|_{[k]} = \pi'\}$ under κ_ν)

$$\mathbb{E}\Big[|\Pi_{(1)}(u)|^{k-1-\gamma}g(|\Pi_{(1)}(u)|)\int_{\mathcal{P}_\mathbb{N}}\kappa_\nu(\mathrm{d}\pi)h(|\pi_i|, 1 \leq i \leq b)\mathbf{1}_{\{\pi|_{[k]}=\pi'\}}\Big],$$

and since f, g, h are compactly supported, the whole integral converges to

$$\int_0^\infty f(u)\mathrm{d}u\mathbb{E}\Big[|\Pi_{(1)}(u)|^{k-1-\gamma}g(|\Pi_{(1)}(u)|)\int_{\mathcal{P}_\mathbb{N}}\kappa_\nu(\mathrm{d}\pi)h(|\pi_i|, 1 \leq i \leq b)\mathbf{1}_{\{\pi|_{[k]}=\pi'\}}\Big],$$

which, by Proposition 6, equals

$$\mathbb{E}\Big[f(D_k)g(|\Pi_{(1)}(D_k-)|)h\Big(\frac{|\Pi_i(D_k)|}{|\Pi_{(1)}(D_k-)|}, 1 \leq i \leq b\Big)\mathbf{1}_{\{[k]\cap\Pi(D_k)=\pi'\}}\Big].$$

It is now easy to conclude, since $|\Pi_i(D_k)| > 0$ almost-surely. $\square$

4.2 Convergence of finite-dimensional marginals

The first step in the proof of Theorem 1 is the following result on reduced trees T_C^B of Section 3.2.4.

Proposition 12. *Let $B \subset \mathbb{N}$ be finite. Under assumption* **(H')**, *we have the following convergence in distribution in $\mathscr{T}_{\mathrm{w}}$:*

$$\frac{1}{a_n} T_{[n]}^B \xrightarrow[n\to\infty]{(d)} \mathcal{R}(\mathcal{T}_{\gamma,\nu}, B)\,.$$

Proof. We use an induction argument on the cardinality of B. For $B = \{i\}$, one can assume by exchangeability (as soon as $n \geq i$) that $B = \{1\}$, and in this case, the reduced tree is $T_{[n]}^B = (\varnothing, D_1^n)$, while $\mathcal{R}(\mathcal{T}_{\gamma,\nu}, \{1\}) = (\varnothing, D_1)$ by Proposition 10. By the second part of Lemma 9, under **(H')**, it holds that

$$\frac{D_1^n}{a_n} \xrightarrow[n\to\infty]{(d)} D_1\,.$$

This initializes the induction. Now, assume that Proposition 12 has been proved for every set B with cardinality at most $k-1$, for some $k \geq 2$. We want to show that the same is true of any set of cardinality $[k]$, and by exchangeability, we may assume that $B = [k]$.

We now recall, using Proposition 11, that conditionally on $D_k^n = \ell$, $[k] \cap \Pi^n(D_k^n) = \pi'$ having $b \geq 2$ blocks and on $\Pi_i^n(D_k^n) = \pi_i, 1 \leq i \leq b$ with respective cardinality $\#\pi_i = n_i$, the tree $T_{[n]}^{[k]}$ has same distribution as

$$\langle \theta^{(i)}, 1 \leq i \leq b \rangle_\ell\,,$$

where $\theta^{(i)}$ has same distribution as $T_{\pi_i}^{\pi_i'}$, and these trees are independent.

The joint distribution of $D_k^n, [k] \cap \Pi^n(D_k^n), (\#\Pi_{(i)}^n(D_k^n), 1 \leq i \leq k)$ is specified by Lemma 8, and its scaling limit by Lemma 10. We obtain by the induction hypothesis that jointly with the above convergence, conditionally on $[k] \cap \Pi^n(D_k^n) = \pi'$,

$$\begin{aligned}\frac{1}{a_n}\theta^{(i)} &= \frac{a_{n_i}}{a_n}\frac{1}{a_{n_i}}\theta^{(i)} \\ &\xrightarrow[n\to\infty]{(d)} |\Pi_i(D_k)|^\gamma \mathcal{T}^{(i)}\,, \qquad 1 \leq i \leq b\,,\end{aligned}$$

where the $\mathcal{T}^{(i)}$ are independent with same laws as $\mathcal{R}(\mathcal{T}_{\gamma,\nu}, \pi_i')$ respectively. Finally, $a_n^{-1} T_{[n]}^{[k]}$ converges to

$$\langle |\Pi_i(D_k)|^\gamma \mathcal{T}^{(i)}, 1 \leq i \leq b \rangle_{D_k}\,,$$

and the $\mathbb{R}$-tree associated with this tree has same distribution as $\mathcal{R}(\mathcal{T}_{\gamma,\nu}, [k])$ by Proposition 10. $\square$

4.3 Tightness in the Gromov-Hausdorff topology

We now want to improve the convergence of Proposition 12 into a convergence of non-reduced trees for the Gromov-Hausdorff topology. Namely

Proposition 13. *Under hypothesis* **(H')**, *we have the convergence in distribution*

$$\frac{1}{a_n} T_n \xrightarrow[n\to\infty]{(d)} \mathcal{T}_{\gamma,\nu}$$

in $\mathscr{T}$, for the Gromov-Hausdorff topology.

This will be proved by first showing a couple of intermediate lemmas.

Lemma 11. *Under assumption* **(H')**, *we have the convergence in distribution:*

$$(\Pi^n(\lfloor a_n t\rfloor), t \geq 0) \xrightarrow[n\to\infty]{(d)} (\Pi(t), t \geq 0)$$

jointly with

$$\Big(\frac{\#\Pi^n_{(i)}(\lfloor a_n t\rfloor)}{n}, t \geq 0\Big) \xrightarrow[n\to\infty]{(d)} (|\Pi_{(i)}(t)|, t \geq 0)$$

for every $i \geq 1$, all these convergences holding jointly.

Proof. The fact that $([k]\cap\Pi^n(\lfloor a_n t\rfloor), t \geq 0)$ converges in the Skorokhod space to $([k]\cap\Pi(t), t \geq 0)$ for every $k \geq 1$ is obtained by using an inductive argument similar to that used in the proof of Proposition 12. We only sketch the argument. The statement is trivial for $k = 1$, so we can assume that $k \geq 2$. The process $[k]\cap\Pi^n(\lfloor a_n\cdot\rfloor)$ remains constant equal to $[k]$ up to time $a_n^{-1}D^n_k$, and jumps to the state $\pi' = [k] \cap \pi$, $\pi = \Pi^n(D^n_k)$. By Lemma 10, $a_n^{-1}D^n_k \to D_k$ as $n \to \infty$, and the latter has a diffuse law by Proposition 6.

After time $a_n^{-1}D^n_k$, given π, the restrictions $\pi'_i \cap \Pi^n(\lfloor a_n\cdot\rfloor + D^n_k)$ have same distribution as $\pi'_i \cap \Pi^{\pi_i}(\lfloor a_n\cdot\rfloor)$, and are independent. By the induction hypothesis and exchangeability, still conditionally on π, this converges to $\pi'_i \cap \Pi^{(i)}(\cdot)$, where $\Pi^{(i)}, i \geq 1$ are i.i.d. copies of Π. Moreover, since the jump times have diffuse laws, two such copies never jump at the same time, from which one concludes that given π, the process $(\pi'_i \cap \Pi^n(\lfloor a_n t\rfloor + D^n_k), 1 \leq i \leq b(\pi'), t \geq 0)$ converges in the Skorokhod space to $(\pi'_i \cap \Pi^{(i)}(t), 1 \leq i \leq b(\pi'), t \geq 0)$. This concludes the inductive step by plugging the initial constancy interval of the process, with length $a_n^{-1}D^n_k$.

The convergence of $\Pi^n(\lfloor a_n\cdot\rfloor)$ in the Skorokhod space follows, because $d_{\mathcal{P}}([k] \cap \pi, \pi) \leq e^{-k}$ for every $\pi \in \mathcal{P}_{\mathbb{N}}$. This shows that $[k] \cap \Pi^n(\lfloor a_n\cdot\rfloor)$ remains uniformly close to $\Pi^n(\lfloor a_n\cdot\rfloor)$.

Next, by Lemma 2, it follows that, jointly with this convergence, for every $i \geq 1$, the finite-dimensional marginals of $(n^{-1}\#\Pi^n_{(i)}(\lfloor a_n t\rfloor), t \geq 0)$ converge in distribution to those of $(|\Pi_{(i)}(t)|, t \geq 0)$, at least for times which are not fixed discontinuity times of the limiting process — the set of such points is always countable, and it turns out that there are none in the present case. Since we also know that the laws of the processes $(n^{-1}\#\Pi^n_{(i)}(\lfloor a_n t\rfloor), t \geq 0)$ are tight when n varies, by Lemma 9 (these processes all have same distribution as $(n^{-1}X_n(\lfloor a_n t\rfloor), t \geq 0)$ by exchangeability), this allows to conclude. $\square$

For $k + 1 \leq i \leq n$, let

$$S^n_i = \inf\{r \geq 0 : [k] \cap \Pi^n_{(i)}(r) = \varnothing\},$$

the first time when the ball indexed i is separated from the k first balls. The random variables $S^n_i, k + 1 \leq i \leq n$ have same distribution by exchangeability. The strong Markov property at the stopping time S^n_i also shows that conditionally on $\Pi^n_{(i)}(S^n_i) = B$, the process $(B\cap\Pi^n(S^n_i+r), r \geq 0)$ has same distribution as Π^B. The tree $\mathtt{t}_{B\cap\Pi^n(S^n_i+\cdot)}$ has thus same distribution as T_B, and can be seen as a subtree of $T_{[n]}$, characterized by the fact that this subtree contains the leaf labeled i, does not contain any of the leaves labeled by an element of $[k]$, and is the maximal subtree of $T_{[n]}$ with this property. In particular, the Gromov-Hausdorff distance between $T^{[k]}_{[n]}$ and $T_{[n]}$ is at most

$$d_{\mathrm{GH}}(T^{[k]}_{[n]}, T_{[n]}) \leq \max_{k+1\leq i\leq n} \mathrm{ht}\left(\mathtt{t}_{\Pi^n_{(i)}(S^n_i)\cap\Pi^n(S^n_i+\cdot)}\right),$$

where $\mathrm{ht}(\mathtt{t})$, called the height of $\mathtt{t}$, is the maximal height of a vertex in $\mathtt{t}$.

Note that if $j \in \Pi^n_{(i)}(S^n_i)$, then $S^n_j = S^n_i$. Therefore, the blocks $\Pi^n_{(i)}(S^n_i), k+1 \leq i \leq n$ are either disjoint or equal. Moreover, the partition π of $[n] \setminus [k]$ with these blocks is clearly exchangeable.

By putting the previous observations together, we obtain by first conditioning on π, and for every $\eta > 0$,

$$\mathbb{P}\left(d_{\mathrm{GH}}(T_{[n]}^{[k]}, T_{[n]}) \geq \eta a_n\right) \leq \mathbb{E}\Bigg[\sum_{i\geq 1} \mathsf{P}^q_{\#\pi_i}(\mathrm{ht} \geq \eta a_n)\Bigg], \qquad \eta > 0. \tag{15}$$

At this point, we need the following uniform estimate for the height of a P_n-distributed tree, which is the key lemma of this section.

Lemma 12. *Assume* **(H')**. *Then for all $p > 0$, there exists a finite constant C_p such that:*

$$\mathsf{P}^q_n\Big(\mathrm{ht} \geq x a_n\Big) \leq \frac{C_p}{x^p}, \quad \forall x > 0, \forall n \geq 1.$$

Before giving the proof of this statement, we end the proof of Proposition 13. Using Lemma 12 for $p = 2/\gamma$ and (15), we obtain

$$\mathbb{P}\left(d_{\mathrm{GH}}(T_{[n]}^{[k]}, T_{[n]}) \geq \eta a_n\right) \leq C_{2/\gamma}\eta^{-2/\gamma}\mathbb{E}\Bigg[\sum_{i\geq 1} \frac{a_{\#\pi_i}^{2/\gamma}}{a_n^{2/\gamma}}\Bigg].$$

By the exchangeability of the partition π of $[n] \setminus [k]$, note that for every measurable function f,

$$\mathbb{E}[f(\#\pi_{(k+1)})] = \frac{1}{n-k}\mathbb{E}\Bigg[\sum_{i=k+1}^{n} f(\#\pi_{(i)})\Bigg] = \mathbb{E}\Bigg[\sum_{i\geq 1}\frac{\#\pi_i}{n-k} f(\#\pi_i)\Bigg].$$

This finally yields

$$\mathbb{P}\left(d_{\mathrm{GH}}(T_{[n]}^{[k]}, T_{[n]}) \geq \eta a_n\right) \leq C_{2/\gamma}\eta^{-2/\gamma}\mathbb{E}\Bigg[\frac{a_{\#\pi_{(k+1)}}^{2/\gamma}(\#\pi_{(k+1)})^{-1}}{a_n^{2/\gamma}n^{-1}}\Bigg].$$

Since the sequence $(a_n^{2/\gamma}n^{-1}, n \geq 1)$ is strictly positive and regularly varying at ∞ with index 1, we get from the Potter's bounds ([11, Th.1.5.6.]) the existence of a finite constant C such that $(a_k^{2/\gamma}k^{-1})/(a_n^{2/\gamma}n^{-1}) \leq C\sqrt{k/n}$ for all $1 \leq k \leq n$. Hence,

$$\mathbb{P}\left(d_{\mathrm{GH}}(T_{[n]}^{[k]}, T_{[n]}) \geq \eta a_n\right) \leq CC_{2/\gamma}\eta^{-2/\gamma}\mathbb{E}\Bigg[\sqrt{\frac{\#\pi_{(k+1)}}{n}}\Bigg].$$

Note that the quantity in the expectation is bounded by 1. By Proposition 11, it holds that $S^n_{k+1}/a_n \to S_{k+1}$ in distribution as $n \to \infty$, where $S_{k+1} = \inf\{t \geq 0 : [k] \cap \Pi_{(k+1)}(t) = \varnothing\}$. This convergence holds jointly with that of $(n^{-1}\#\Pi^n_{(k+1)}(\lfloor a_n t\rfloor),\ t \geq 0)$ to $(|\Pi_{(k+1)}(t)|, t \geq 0)$ in the Skorokhod space, whence we deduce that

$$\limsup_{n\to\infty}\mathbb{P}\left(d_{\mathrm{GH}}(T_{[n]}^{[k]}, T_{[n]}) \geq \eta a_n\right) \leq CC_{2/\gamma}\eta^{-2/\gamma}\mathbb{E}\left[\sqrt{|\Pi_{(k+1)}(S_{k+1}-)|}\right].$$

Let $S'_{k+1} = \inf\{t \geq 0 : \{2, 3, \dots, k+1\} \cap \Pi_{(1)}(t) = \varnothing\}$, then by exchangeability,

$$\limsup_{n\to\infty}\mathbb{P}\left(d_{\mathrm{GH}}(T_{[n]}^{[k]}, T_{[n]}) \geq \eta a_n\right) \leq CC_{2/\gamma}\eta^{-2/\gamma}\mathbb{E}\left[\sqrt{|\Pi_{(1)}(S'_{k+1}-)|}\right].$$

Since the quantity in the expectation goes to 0 a.s. as $k \to \infty$ and is bounded (indeed $S'_k \uparrow D_{\{1\}}$ a.s. and $\Pi_{(1)}(D_{\{1\}}-) = 0$ by (2) and Proposition 7), we conclude that for every $\eta > 0$,

$$\lim_{k\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(d_{\mathrm{GH}}(T_{[n]}^{[k]}, T_{[n]}) \geq \eta a_n\right) = 0. \tag{16}$$

It is now easy to conclude from this, Proposition 4.2 and the fact that $\mathcal{R}(\mathcal{T}_{\gamma,\nu}, [k])$ converges is distribution in $(\mathscr{T}, d_{\mathrm{GH}})$ to $\mathcal{T}_{\gamma,\nu}$ as $k \to \infty$ ([24]), using [10, Theorem 4.2]:

Lemma 13. *Let X_n, X, X_n^k, X^k be random variables in a metric space (M, d). We assume that for every k, we have $X_n^k \to X^k$ in distribution as $n \to \infty$, and $X^k \to X$ in distribution as $k \to \infty$. Finally, we assume that for every $\eta > 0$,*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(d(X_n^k, X_n) > \eta) = 0\,.$$

Then $X_n \to X$ in distribution as $n \to \infty$.

Proof of Lemma 12. Note that if the statement holds for some $p > 0$, it then holds for all $q \in (0, p)$. We can therefore assume in the following that $p > 1/\gamma$ and we let $\varepsilon > 0$ be so that $p(\gamma - \varepsilon) > 1$. The main idea of the proof is to proceed by induction on n, using the Markov branching property. We start with some technical preliminaries.

• First note that $\mathsf{E}_n^q[\mathrm{ht}^r] < \infty$ for all $r > 0$ and all $n \geq 1$. This can easily be proved by induction on n (r being fixed) using the Markov branching property, and the facts that $q_n((n)) < 1$ and that $H_1 = 1$ almost-surely.

• Second, we replace the sequence $(a_n, n \geq 1)$ by a 'nicer' sequence $(\tilde{a}_n, n \geq 1)$ such that $\tilde{a}_n \sim a_n$, i.e. $a_n/\tilde{a}_n \to 1$ as $n \to \infty$ (this step is trivial when $a_n = n^\gamma$; we then take $\tilde{a}_n = a_n$). Since $(a_n, n \geq 1)$ is regularly varying at ∞, it is well-known (see [11, Theorem 1.3.1]) that it can be written in the form

$$a_n = c(n) \exp\left(\int_1^n \varepsilon(u)\mathrm{d}u/u\right), \quad n \geq 1,$$

where $c(n) \to c > 0$ as $n \to \infty$ and ε is a measurable function that converges to 0 as ∞. Define

$$\tilde{a}_n = c \exp\left(\int_1^n \varepsilon(u)\mathrm{d}u/u\right), \quad n \geq 1.$$

We claim that there exists an integer $n_\varepsilon \geq 1$ such that for $n \geq n_\varepsilon$,

$$\frac{\tilde{a}_k}{\tilde{a}_n} \leq \left(\frac{k}{n}\right)^{\gamma-\varepsilon} \qquad \forall 1 \leq k \leq n. \tag{17}$$

Indeed, let u_ε be such that $|\varepsilon(u)| \leq \varepsilon$ for all $u \geq u_\varepsilon$. For $n \geq k \geq u_\varepsilon$, we have

$$\left|\int_k^n -\varepsilon(u)\mathrm{d}u/u\right| \leq \varepsilon \int_k^n \mathrm{d}u/u \leq \varepsilon \ln(n/k)$$

hence

$$\frac{\tilde{a}_k}{\tilde{a}_n} = \left(\frac{k}{n}\right)^\gamma \exp\left(-\int_k^n \varepsilon(u)\mathrm{d}u/u\right) \leq \left(\frac{k}{n}\right)^{\gamma-\varepsilon}.$$

Besides $\sup_{k\in\{1,\ldots,\lfloor u_\varepsilon\rfloor\}} \tilde{a}_k k^{\varepsilon-\gamma}/(\tilde{a}_n n^{\varepsilon-\gamma}) \leq 1$ for all n large enough (say $n \geq n'_\varepsilon$). Hence $\tilde{a}_k/\tilde{a}_n \leq (k/n)^{\gamma-\varepsilon}$ for all $n \geq n_\varepsilon = \max(n'_\varepsilon, u_\varepsilon)$ and all $1 \leq k \leq n$.

• Since $\tilde{a}_n > 0$ for all $n \geq 1$ and $\tilde{a}_n \sim a_n$, there exists some $c > 0$ such that $a_n \geq c\tilde{a}_n$ for all $n \geq 1$. It is therefore sufficient to prove the existence of a finite C_p (a priori different from the one in the statement of the lemma) such that

$$\mathsf{P}_n^q(\mathrm{ht} \geq x\tilde{a}_n) \leq \frac{C_p}{x^p}, \quad \forall x > 0 \text{ and } n \geq 1, \tag{18}$$

to finish the proof of the lemma. In order to prove (18), we will use the integer n_ε introduced around (17) and we will further assume, taking n_ε larger if necessary, that $\tilde{a}_n \geq 1$ for every $n \geq n_\varepsilon$. Introduce now $0 < C_p^1 < 1$ such that

$$(1-u)^{-p} \leq 1 + 2pu, \quad \forall 0 \leq u \leq C_p^1.$$

Using **(H')** and the fact that $q_n((n)) < 1$ for all $n \geq 1$, there exists also $C_p^2 > 0$ such that

$$\tilde{a}_n \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) \left(1 - \sum_{i=1}^{p(\lambda)} \left(\frac{\lambda_i}{n} \right)^{(\gamma-\varepsilon)p} \right) \geq C_p^2, \quad \forall n \geq 1 \tag{19}$$

(recall that $(\gamma - \varepsilon)p > 1$ and $\tilde{a}_n > 0$ for all $n \geq 1$). Last we let

$$C_p(n_\varepsilon) := \max_{1 \leq n \leq n_\varepsilon} \left(\mathsf{E}_n^q[\mathrm{ht}^p]/\tilde{a}_n^p \right) < \infty$$

and we set

$$C_p := \max \left(C_p(n_\varepsilon), \left(1/C_p^1, 2p/C_p^2 \right)^p \right) < \infty.$$

Our goal is to prove by induction on $n \geq 1$ that

$$\mathsf{P}_n^q(\mathrm{ht} < x\tilde{a}_n) \geq 1 - \frac{C_p}{x^p}, \qquad \text{for every } x > 0, \tag{A$_n$}$$

Clearly, (A_n) holds for all $n \leq n_\varepsilon$ since $C_p \geq C_p(n_\varepsilon)$ and $\mathsf{P}_n^q(\mathrm{ht} \geq x\tilde{a}_n) \leq \mathsf{E}_n^q[\mathrm{ht}^p]/(x\tilde{a}_n)^p$. Now assume that (A_k) is satisfied for all $k \leq n-1$ for some $n \geq n_\varepsilon$. For all $0 < x \leq C_p^{1/p}$, the expected inequalities in (A_n) are obvious, so it remains to prove them for $x > C_p^{1/p}$. To get (A_n), we will prove by induction on $i \in \mathbb{N}$ that

$$\mathsf{P}_n^q(\mathrm{ht} < x\tilde{a}_n) \geq 1 - \frac{C_p}{x^p}, \qquad \text{for every } x \in \left(0, \frac{i}{\tilde{a}_n} \right), \tag{A$_{n,i}$}$$

which will obviously lead to (A_n). Note first that $(\mathrm{A}_{n,1})$ holds since $1/\tilde{a}_n \leq 1 \leq C_p^{1/p}$. Assume next that $(\mathrm{A}_{n,i})$ is true, and fix $x \in (0, (i+1)/\tilde{a}_n)$. Using the Markov branching property and the fact that (A_k), holds for every $k \leq n-1$, as well as $(\mathrm{A}_{n,i})$, we get

$$\begin{aligned}
\mathsf{P}_n^q(\mathrm{ht} < x\tilde{a}_n) &= \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) \prod_{i=1}^{p(\lambda)} \mathsf{P}_{\lambda_i}^q(\mathrm{ht} < x\tilde{a}_n - 1) \\
&\geq \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) \prod_{i=1}^{p(\lambda)} \left(1 - \frac{C_p \tilde{a}_{\lambda_i}^p}{(x\tilde{a}_n - 1)^p} \right)^+ \\
&\geq \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) \left(1 - \sum_{i=1}^{p(\lambda)} \frac{C_p \tilde{a}_{\lambda_i}^p}{(x\tilde{a}_n - 1)^p} \right),
\end{aligned}$$

using the notation $r^+ = \max(r, 0)$ and that for all sequences of non-negative terms $b_i, i \geq 1$, $\prod_{i=1}^m (1 - b_i)^+ \geq 1 - \sum_{i=1}^m b_i$, for every $m \geq 1$. We can assume that $x > C_p^{1/p}$ since (A_n) holds otherwise. In particular, $x\tilde{a}_n \geq x > 1/C_p^1 > 1$. Therefore,

$$\frac{1}{(x\tilde{a}_n - 1)^p} = \frac{1}{(x\tilde{a}_n)^p (1 - 1/(x\tilde{a}_n))^p} \leq \frac{1 + 2p/(x\tilde{a}_n)}{(x\tilde{a}_n)^p},$$

and then

$$\begin{aligned}
&\mathsf{P}_n^q(\mathrm{ht} < x\tilde{a}_n) \\
\geq\ & \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) - \frac{C_p}{x^p} \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) \sum_{i=1}^{p(\lambda)} \left(\frac{\tilde{a}_{\lambda_i}}{\tilde{a}_n} \right)^p - \frac{2pC_p}{x^{p+1}\tilde{a}_n} \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) \sum_{i=1}^{p(\lambda)} \left(\frac{\tilde{a}_{\lambda_i}}{\tilde{a}_n} \right)^p \\
\underset{\text{by (17)}}{\geq}\ & 1 - \frac{C_p}{x^p} \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) \sum_{i=1}^{p(\lambda)} \left(\frac{\lambda_i}{n} \right)^{(\gamma-\varepsilon)p} - \frac{2pC_p}{x^{p+1}\tilde{a}_n} \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) \sum_{i=1}^{p(\lambda)} \left(\frac{\lambda_i}{n} \right)^{(\gamma-\varepsilon)p} \\
\geq\ & 1 - \frac{C_p}{x^p} + \frac{C_p}{x^p} \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) \left(1 - \sum_{i=1}^{p(\lambda)} \left(\frac{\lambda_i}{n} \right)^{(\gamma-\varepsilon)p} \right) - \frac{2pC_p}{x^{p+1}\tilde{a}_n} \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) \sum_{i=1}^{p(\lambda)} \left(\frac{\lambda_i}{n} \right)^{(\gamma-\varepsilon)p}.
\end{aligned}$$

We then use (19) and the fact that $\sum_{i=1}^{p(\lambda)} (\lambda_i/n)^{(\gamma-\varepsilon)p} \le 1$ (since $(\gamma-\varepsilon)p > 1$) to get

$$\mathsf{P}_n^q(\mathrm{ht} < x\tilde{a}_n) \ge 1 - \frac{C_p}{x^p} + \frac{C_p}{x^p \tilde{a}_n}\left(C_p^2 - \frac{2p}{x}\right).$$

By assumption, $x > C_p^{1/p} \ge 2p/C_p^2$, hence

$$\mathsf{P}_n^q(\mathrm{ht} < x\tilde{a}_n) \ge 1 - \frac{C_p}{x^p} \qquad \text{for every } x \in \left(0, \frac{i+1}{\tilde{a}_n}\right),$$

as wanted. □

4.4 Incorporating the measure

We now finish the proof of Theorem 1, by improving the Gromov-Hausdorff convergence of Proposition 13 to a Gromov-Hausdorff-Prokhorov convergence, when the uniform measure $\mu_n = \mu_{\partial T_n}$ on leaves is added to T_n in order to view it as an element of $\mathscr{T}_{\mathrm{w}}$ rather than $\mathscr{T}$.

We will use the fact [22, Lemma 2.3] that the convergence in distribution of $a_n^{-1}T_n$ as $n \to \infty$ in $\mathscr{T}$ entails that the laws of the random variables $a_n^{-1}T_n$, form a tight sequence of probability measures on $\mathscr{T}_{\mathrm{w}}$. Therefore, it suffices to identify the limit as $\mathcal{T}_{\gamma,\nu}$.

So let us assume that $a_n^{-1}T_n$ converges to $(\mathcal{T}', d', \rho', \mu') \in \mathscr{T}_{\mathrm{w}}$ in distribution, when $n \to \infty$ along some subsequence. Let $L_1^n, L_2^n, \ldots, L_k^n$ be k i.i.d. uniform leaves of T_n. Conditionally given the event that these leaves are pairwise distinct, which occurs with probability going to 1 as $n \to \infty$ with k fixed, these leaves are just a uniform sample of k distinct leaves of T_n, so by Lemma 6 and exchangeability, the subtree of T_n spanned by the root and the leaves $L_1^n, \ldots, L_k^n$ has same distribution as $T_{[n]}^{[k]}$. By Proposition 12, we know that $a_n^{-1}T_{[n]}^{[k]}$ converges in distribution to $\mathcal{R}(\mathcal{T}_{\gamma,\nu}, [k])$ in $\mathscr{T}$.

A k-pointed compact metric space is an object of the form $((X,d), x_1, \ldots, x_k)$ where (X,d) is a compact metric space and $x_1, \ldots, x_k \in X$. The set of k-pointed metric spaces can be endowed with the k-pointed Gromov-Hausdorff distance

$$d_{\mathrm{GH}}^{(k)}\Big(((X,d), x_1, \ldots, x_k), ((X',d'), x_1', \ldots, x_k')\Big) = \inf_{\phi,\phi'} \max_{1\le i\le k} \mathrm{dist}(\phi(x_i), \phi'(x_i')) \vee \mathrm{dist}_{\mathrm{H}}(\phi(X), \phi'(X')),$$

where, as in the definition of the Gromov-Hausdorff distance, the infimum is over isometric embeddings ϕ, ϕ' of X, X' into some common space (M, dist). Note in particular that $d_{\mathrm{GH}}^{(1)} = d_{\mathrm{GH}}$. Now, the fact that $a_n^{-1}T_n$ converges to $(\mathcal{T}', d', \rho', \mu')$ in $\mathscr{T}_{\mathrm{w}}$ implies that the $k+1$-pointed space $(a_n^{-1}T_n, \rho, L_1^n, \ldots, L_k^n)$ converges in distribution to $(\mathcal{T}', \rho', L_1, \ldots, L_k)$, where $L_1, \ldots, L_k$ are i.i.d. with law μ' conditionally on the latter. See [30, Proposition 10] for a proof and further properties of the k-pointed Gromov-Hausdorff distance, which in particular is Polish.

If (T, d, ρ) is a rooted $\mathbb{R}$-tree and $x_1, \ldots, x_k \in T$, the union of geodesics from ρ to the x_i's

$$R(T, x_1, \ldots, x_k) = \bigcup_{i=1}^{k} [[\rho, x_i]]$$

is in turn an $\mathbb{R}$-tree rooted at ρ with at most k leaves, called the subtree of T spanned by $x_1, \ldots, x_k$ (the role of the root being implicit).

Lemma 14. *Let $(\mathcal{A}_n, d_n, \rho_n), n \ge 1$ be a sequence of rooted $\mathbb{R}$-trees and $x_1^n, \ldots, x_k^n$ be k points in $\mathcal{A}_n$, such that $((\mathcal{A}_n, d_n), \rho_n, x_1^n, \ldots, x_n^k)$ converges for the $k+1$-pointed Gromov-Hausdorff distance to a limit $((\mathcal{A}, d), \rho, x_1, \ldots, x_k)$. Then the subtree $R(\mathcal{A}_n, x_1^n, \ldots, x_k^n)$ converges in $\mathscr{T}$ to the subtree $R(\mathcal{A}, x_1, \ldots, x_k)$.*

We will prove this lemma at the end of the section. By using the Skorokhod representation theorem, we may assume that the convergence of $(a_n^{-1}T_n, \rho, L_1^n, \ldots, L_k^n)$ to $(\mathcal{T}', \rho', L_1, \ldots, L_k)$ holds almost-surely. This, together with Lemma 14 and the discussion at the beginning of this section, implies the joint convergence in distribution in $\mathscr{T}$ of $a_n^{-1}T_{[n]}, a_n^{-1}T_{[n]}^{[k]}$ to $\mathcal{T}', \mathcal{T}_k'$, still along the appropriate subsequence, where $\mathcal{T}_k'$ is the subtree of $\mathcal{T}'$ spanned by $L_1, \ldots, L_k$. In particular, this identifies the law of $\mathcal{T}_k'$ as that of $\mathcal{R}(\mathcal{T}_{\gamma,\nu}, [k])$. When $k \to \infty$, we already stressed that the latter trees converge (in distribution in $\mathscr{T}_{\mathrm{w}}$, with the uniform measure μ_k on the set of its k leaves) to $\mathcal{T}_{\gamma,\nu}$. On the other hand, $\mathcal{T}_k'$ converges a.s. to $(\mathcal{T}'', d', \rho', \mu')$ in $\mathscr{T}_{\mathrm{w}}$ as $k \to \infty$, where $\mathcal{T}''$ is the closure in $\mathcal{T}'$ of

$$\bigcup_{i=1}^{\infty} [[\rho', L_i]] \, .$$

But the joint convergence of $T_{[n]}, T_{[n]}^{[k]}$ in $\mathscr{T}$ along some subsequence and (16) imply that for every $\eta > 0$, $\lim_{k\to\infty} \mathbb{P}(d_{\mathrm{GH}}(\mathcal{T}_k', \mathcal{T}') > \eta) = 0$. So $\mathcal{T}'' = \mathcal{T}'$ a.s., entailing that $(\mathcal{T}', d', \rho', \mu')$ has same law as $\mathcal{T}_{\gamma,\nu}$. This identifies the limit of $a_n^{-1}T_n$ in $\mathscr{T}_{\mathrm{w}}$ as $\mathcal{T}_{\gamma,\nu}$, ending the proof of Theorem 1.

It remains to prove Lemma 14. We only sketch the argument, leaving the details to the reader. We use induction on k. For $k = 1$, the subtree $R(\mathcal{A}_n, x_1^n)$ is isometric to a real segment $[0, d_n(\rho_n, x_1^n)]$ rooted at 0. The 2-pointed convergence of $((\mathcal{A}_n, d_n), \rho_n, x_1^n)$ to $((\mathcal{A}, d), \rho, x_1)$ entails that $d_n(\rho_n, x_1^n)$ converges to $d(\rho, x_1)$, hence that $R(\mathcal{A}_n, x_1^n)$ converges to $[0, d(\rho, x_1)]$ rooted at 0, which is isometric to $R(\mathcal{A}, x_1)$.

For the induction step, we use the general fact that if $\mathcal{A}$ is a rooted $\mathbb{R}$-tree and $x_1, \ldots, x_k, x_{k+1} \in \mathcal{A}$, then the distance between x_{k+1} and the subtree of $\mathcal{A}$ spanned by $x_1, \ldots, x_k$ is equal to

$$\delta_{k+1} = \min_{1 \le i \le k} \left(\frac{d(x_{k+1}, x_i) + d(x_{k+1}, \rho) - d(x_i, \rho)}{2} \right) .$$

Moreover, if $i \in \{1, 2, \ldots, k\}$ is an index that realizes this minimum, then the branchpoint $x_{k+1} \wedge x_i$ is at distance δ_{k+1} from x_{k+1} and is the ancestor of x_i at height (i.e. distance from ρ)

$$h_{k+1} = d(\rho, x_{k+1}) - \delta_{k+1} \, .$$

Roughly speaking, we get that $R(\mathcal{A}, x_1, \ldots, x_{k+1})$ is obtained from $R(\mathcal{A}, x_1, \ldots, x_k)$ by grafting a segment with length δ_{k+1} at the ancestor of x_i with height h_{k+1}.

In our particular situation, and with obvious notations, we get that $R(\mathcal{A}_n, x_1^n, \ldots, x_{k+1}^n)$ is obtained by grafting a segment with length δ_{k+1}^n to the ancestor of $x_{i_n}^n$ with height h_{k+1}^n, where i_n is some index in $\{1, \ldots, k\}$ that can depend on n. Up to extracting, we may assume that $i_n = i$ is constant. The $k+2$-pointed convergence of $((\mathcal{A}_n, d_n), \rho_n, x_1^n, \ldots, x_{k+1}^n)$ to $((calA, d), \rho, x_1, \ldots, x_{k+1})$ entails that the d_n-distances between elements of $\{\rho_n, x_1^n, \ldots, x_{k+1}^n\}$ converge to the corresponding d-distances of elements in $\{\rho, x_1, \ldots, x_k\}$. Consequently, it holds that $\delta_{k+1}^n, h_{k+1}^n$ converge to δ_{k+1}, h_{k+1}, defined as above. Together with the induction hypothesis stating that $R(\mathcal{A}_n, x_1^n, \ldots, x_k^n)$ converges in $\mathscr{T}$ to $R(\mathcal{A}, x_1, \ldots, x_k)$, this entails easily that $R(\mathcal{A}_n, x_1^n, \ldots, x_{k+1}^n)$ converges in $\mathscr{T}$ to the $\mathbb{R}$-tree obtained by grafting a segment with length δ_{k+1} to the ancestor of x_i with height h_{k+1} in $R(\mathcal{A}, x_1, \ldots, x_k)$, and this tree is $R(\mathcal{A}, x_1, \ldots, x_{k+1})$. The result being independent of the particular value of i (selected by the choice of an extraction), the convergence holds without taking extractions, which concludes the proof.

4.5 Proof of Theorem 2

To pass from trees with n vertices (with law Q_n^q) to trees with laws of the form $\mathsf{P}_n^{q'}$, with n leaves, we introduce a transformation on trees, in which every vertex which is not a leaf is attached to an extra "ghost" neighbor, which is a leaf.

Precisely, if $\mathfrak{t}$ is a plane tree, then the modification $\mathfrak{t}^\circ$ is defined as

$$\mathfrak{t}^\circ = \mathfrak{t} \cup \bigcup_{u=(u_1,\dots,u_k)\in\mathfrak{t}\setminus\partial\mathfrak{t}} \{(u_1,\dots,u_k,c_u(\mathfrak{t})+1)\}\,.$$

If we are given a tree rather than a plane tree, then this construction performed on any plane representative a the tree $\mathtt{t}$ will yield plane trees in the same equivalence class, which we call $\mathtt{t}^\circ$. Note that

$$\#\partial\mathtt{t}^\circ = \#\mathtt{t}\,.$$

We see $\mathtt{t}^\circ$ as an element of $\mathscr{M}_{\mathrm{w}}$ (endowed with graph distance and uniform distribution on $\partial\mathtt{t}^\circ$), and view $\mathtt{t}$ as an element of $\mathscr{M}_{\mathrm{w}}$, by endowing it also with the graph distance, but this time, with the uniform distribution $\mu_{\mathtt{t}}$ on $\mathtt{t}$. It is easy to see, using the natural isometric embedding of $\mathtt{t}$ into $\mathtt{t}^\circ$, that for every $a>0$,

$$d_{\mathrm{GHP}}(a\mathtt{t},a\mathtt{t}^\circ)\leq a\,. \tag{20}$$

Let $(q_n,n\geq 1)$ be, as in Section 1.2.2, a family of probability distributions respectively on $\mathcal{P}_n$, such that $q_1((1))=1$. We introduce the family $q_n^\circ, n\geq 1$ of probability measures respectively on $\mathcal{P}_n$ by $q_1^\circ(\emptyset)=1$, and

$$q_{n+1}^\circ((\lambda,1)) = q_n(\lambda)\,,\qquad n\geq 1,\lambda\in\mathcal{P}_n\,,$$

where $(\lambda,1)=(\lambda_1,\dots,\lambda_{p(\lambda)},1)\in\mathcal{P}_{n+1}$.

It is then immediate to show by induction that if T_n has law Q_n^q, then T_n° has law $\mathsf{P}_n^{q^\circ}$, with the notation of Section 1.2.1. We leave this verification to the reader. In view of this and (20), we see that Theorem 2 is a straightforward consequence of the following statement.

Lemma 15. *If $(q_n,n\geq 1)$ satisfies* **(H)** *with either $\gamma\in(0,1)$, or $\gamma=1$ and $\ell(n)\to 0$ as $n\to\infty$, then $(q_n^\circ,n\geq 1)$ satisfies* **(H)***, with same fragmentation pair $(-\gamma,\nu)$ and function ℓ.*

Proof. Let $f:\mathcal{S}^\downarrow\to\mathbb{R}$ be a Lipschitz function with uniform norm and Lipschitz constant bounded by K. Let also $g(\mathbf{s})=(1-s_1)f(\mathbf{s})$. Then

$$\left|f\Big(\frac{(\lambda,1)}{n+1}\Big)-f\Big(\frac{\lambda}{n}\Big)\right|\leq K\sum_{i=1}^{p(\lambda)}\frac{\lambda_i}{n(n+1)}+\frac{K}{n+1}\leq\frac{2K}{n+1}\,,$$

so that

$$\begin{aligned}|\overline{q}_{n+1}^\circ(g)-\overline{q}_n(g)| &\leq \sum_{\lambda\in\mathcal{P}_n} q_n(\lambda)\left|\Big(1-\frac{\lambda_1}{n+1}\Big)f\Big(\frac{(\lambda,1)}{n+1}\Big)-\Big(1-\frac{\lambda_1}{n}\Big)f\Big(\frac{\lambda}{n}\Big)\right| \\ &\leq \sum_{\lambda\in\mathcal{P}_n} q_n(\lambda)\left(\frac{K\lambda_1}{n(n+1)}+\frac{2K}{n+1}\right) \\ &\leq \frac{3K}{n+1}\end{aligned}$$

multiplying both sides by $n^\gamma\ell(n)$, we see that the upper-bound converges to 0 as $n\to\infty$ under our hypotheses. Since $n^\gamma\ell(n)\overline{q}_n(g)$ converges to $\nu(g)$ by **(H)**, we obtain the same convergence with $\overline{q}_n^\circ$ instead of $\overline{q}_n$. This yields the result. □

4.6 Proof of Proposition 4

Recall the notation $\Lambda^{(\mathbf{s})}(n)$ for the decreasing sequence of sizes of blocks restricted to $\{1,\dots,n\}$ of a random variable with painbox distribution $\rho_{\mathbf{s}}(\mathrm{d}\pi)$, with $\mathbf{s}\in\mathcal{S}^\downarrow$, $\sum_{i\geq1}s_i=1$. Recall also that

$\Lambda^{(\mathbf{s})}(n)/n \to \mathbf{s}$ in $\mathcal{S}^{\downarrow}$ almost-surely. Now set for $\lambda \in \mathcal{P}_n$,

$$\begin{aligned}\tilde{q}_n(\lambda) &= n^{-\gamma} \int_{\mathcal{S}^{\downarrow}} \mathbb{P}(\Lambda^{(\mathbf{s})}(n) = \lambda) \mathbf{1}_{\{n^{-\gamma/2} \leq 1 - s_1\}} \nu(\mathrm{d}\mathbf{s}), \quad \lambda \neq (n), \\ \tilde{q}_n((n)) &= 1 - \sum_{\lambda \in \mathcal{P}_n, \lambda \neq (n)} \tilde{q}_n(\lambda).\end{aligned}$$

For n large enough, say $n \geq n_0$,

$$0 < \sum_{\lambda \in \mathcal{P}_n, \lambda \neq (n)} \tilde{q}_n(\lambda) \leq n^{-\gamma/2} \int_{\mathcal{S}^{\downarrow}} (1 - s_1) \nu(\mathrm{d}\mathbf{s}) \leq 1,$$

hence $\tilde{q}_n$ defines a probability distribution on $\mathcal{P}_n$ such that $\tilde{q}_n((n)) < 1$. Then set $q_n = \tilde{q}_n$ for $n \geq n_0$ and for $n < n_0$ let q_n be any distribution on $\mathcal{P}_n$ such that $q_n((n)) < 1$.

Next, consider a continuous function $f : \mathcal{S}^{\downarrow} \to \mathbb{R}_+$. For $n \geq n_0$, we have

$$n^{\gamma} \sum_{\lambda \in \mathcal{P}_n} q_n(\lambda) \left(1 - \frac{\lambda_1}{n}\right) f\left(\frac{\lambda}{n}\right) = \int_{\mathcal{S}^{\downarrow}} \mathbb{E}\left[\left(1 - \frac{\Lambda_1^{(\mathbf{s})}(n)}{n}\right) f\left(\frac{\Lambda^{(\mathbf{s})}(n)}{n}\right)\right] \mathbf{1}_{\{n^{-\gamma/2} \leq 1 - s_1\}} \nu(\mathrm{d}\mathbf{s}),$$

which converges to $\int_{\mathcal{S}^{\downarrow}} f(\mathbf{s})(1 - s_1)\nu(\mathrm{d}\mathbf{s})$ as $n \to \infty$ by dominated convergence. This concludes the proof.

5 Scaling limits of conditioned Galton-Watson trees

Recall the notations of Section 2.1. While the probability distribution GW_ξ enjoys the so-called branching property, it holds that the conditioned versions $\mathrm{GW}_\xi^{(n)}$ are Markov branching trees.

Proposition 14. (i) *One has* $\mathrm{GW}_\xi^{(n)} = \mathsf{Q}_n^q$ *for every* $n \geq 1$, *where the splitting probabilities* $q = (q_n, n \geq 1)$ *are defined by* $q_1((1)) = 1$ *and for every* $n \geq 2$ *and* $\lambda = (\lambda_1, \ldots, \lambda_p) \in \mathcal{P}_n$,

$$q_n(\lambda) = \frac{p!}{\prod_{j \geq 1} m_j(\lambda)!} \xi(p) \frac{\prod_{i=1}^p \mathrm{GW}_\xi(\#\mathbb{t} = \lambda_i)}{\mathrm{GW}_\xi(\#\mathbb{t} = n+1)} \,. \tag{21}$$

(ii) *On some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, *let* $X_1, X_2, \ldots$ *be i.i.d. with distribution* $\mathbb{P}(X_1 = k) = \mathrm{GW}_\xi(\#\mathbb{t} = k)$, *and set* $\tau_p = X_1 + X_2 + \ldots + X_p$. *Then*

$$q_n(p(\lambda) = p) = \xi(p) \frac{\mathbb{P}(\tau_p = n)}{\mathbb{P}(\tau_1 = n+1)} \,,$$

and $q_n(\cdot \,|\, \{p(\lambda) = p\})$ *is the law of the non-increasing rearrangement of* $(X_1, \ldots, X_p)$ *conditionally on* $X_1 + \ldots + X_p = n$.

Proof. (i) Under GW_ξ (viewed as a law on plane trees), conditionally on $c_\varnothing = p$, the p (plane) subtrees born from $\varnothing$ are independent with law GW_ξ. For integers $a_1, \ldots, a_p$ with sum n, the probability that these trees have sizes equal to $a_1, \ldots, a_p$ is thus $\prod_{i=1}^p \mathrm{GW}_\xi(\#\mathbb{t} = a_i)$. Hence,

$$\mathrm{GW}_\xi^{(n+1)}(c_\varnothing = p, \#\mathbb{t}_i = a_i, 1 \leq i \leq p) = \xi(p) \frac{\prod_{i=1}^p \mathrm{GW}_\xi(\#\mathbb{t} = a_i)}{\mathrm{GW}_\xi(\#\mathbb{t} = n+1)} \,, \tag{22}$$

and conditionally on the event on the left-hand side, the subtrees born from the root are independent with respective laws $\mathrm{GW}_\xi^{(a_i)}, 1 \leq i \leq p$. Letting λ be the non-increasing rearrangement of $(a_1, \ldots, a_p)$ and re-ordering the subtrees by non-increasing order of size (with some convention for

ties, e.g. taking them in order of appearance according to the plane structure), we see that these subtrees are independent with laws $\mathrm{GW}_\xi^{(\lambda_i)}, 1 \leq i \leq p$. Using the fact that there are $p!/\prod_{j\geq 1} m_j(\lambda)!$ compositions $(a_1, \ldots, a_p)$ of the integer n corresponding to a partition $\lambda \in \mathcal{P}_n$, and viewing GW_ξ as a law on T instead of plane trees, the conclusion easily follows.

(ii) We have $q_n(p(\lambda) = p) = \mathrm{GW}_\xi^{(n+1)}(c_\varnothing = p)$, and the wanted result is just an interpretation of (22). $\square$

On the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as in the previous statement, we will also assume that $(S_r, r \geq 0)$ is a random walk with i.i.d. steps, each having distribution $\xi(i+1), i \geq -1$. Then the well-known Otter-Dwass formula (or *cyclic lemma*) [32, Chapter 6] allows to rewrite

$$q_n(p(\lambda) = p) = \xi(p)\frac{\frac{p}{n}\mathbb{P}(S_n = -p)}{\frac{1}{n+1}\mathbb{P}(S_{n+1} = -1)} = \frac{n+1}{n}\hat{\xi}(p)\frac{\mathbb{P}(S_n = -p)}{\mathbb{P}(S_{n+1} = -1)}\,, \tag{23}$$

where $\hat{\xi}(p) = p\xi(p)$ is the size-biased distribution associated with ξ.

It is often convenient to work with size-biased orderings of the sequence $(X_1, \ldots, X_p)$ rather than with its non-increasing rearrangement. Recall that if $(x_1, x_2, \ldots)$ is a non-negative sequence with $\sum_i x_i < \infty$, we define its size-biased ordering in the following way. If all terms are zero then we let $x_1^* = 0$, otherwise we let i^* be a random variable with

$$\mathbb{P}(i^* = i) = \frac{x_i}{\sum_{j\geq 1} x_j}\,,$$

and set $x_1^* = x_{i^*}$. We then remove the i^*-th term from the sequence $(x_i, i \geq 1)$ and resume the procedure, defining a random re-ordering $(x_1^*, x_2^*, \ldots)$ of the sequence $(x_1, x_2, \ldots)$. The size-biased ordering $(X_1^*, X_2^*, \ldots)$ of a random sequence $(X_1, X_2, \ldots)$ is defined similarly, by first conditioning on $(X_1, X_2, \ldots)$. If μ is the law of $(X_1, X_2, \ldots)$, we let μ^* be the law of $(X_1^*, X_2^*, \ldots)$.

If μ is a probability distribution on $\mathcal{S}^\downarrow$, then μ^* is a probability distribution on the set $\mathcal{S}_1 = \{\mathbf{x} = (x_1, x_2 \ldots) \in [0,1]^{\mathbb{N}} : \sum_{i\geq 1} x_i \leq 1\}$ which is endowed with any metric inducing the product topology — in particular, $\mathcal{S}_1$ is compact. Similarly, if μ is a non-negative measure on $\mathcal{S}^\downarrow$, we let $\mu^*(f) = \int_{\mathcal{S}^\downarrow} \mu(\mathrm{d}\mathbf{s})\mathbb{E}[f(\mathbf{s}^*)]$, for every non-negative measurable $f : \mathcal{S}_1 \to \mathbb{R}_+$, where $\mathbf{s}^*$ is the size-biased reordering of $\mathbf{s}$. The following statement is a simple variation of [9, Proposition 2.3], replacing probability distributions with finite measures.

Lemma 16. *Let $\mu_n, n \geq 1$ and μ be finite measures on $\mathcal{S}^\downarrow$, and assume that μ is supported on $\{\mathbf{s} \in \mathcal{S}^\downarrow : \sum_i s_i = 1\}$. Then μ_n converges weakly to μ if and only if μ_n^* converges weakly to μ^*.*

5.1 Finite variance case

Here we assume that ξ has finite variance $\sum_{p\geq 1} p(p-1)\xi(p) = \sigma^2 < \infty$. In the proofs to come, C will denote a positive, finite constant with values that can differ from line to line.

Our goal is to check hypothesis **(H)** for the sequence q of (21), and for the measure $\nu = (\sigma/2)\nu_2$. Due to Lemma 16, it suffices to show that

$$n^{1/2}((1 - s_1)\overline{q}_n(\mathrm{d}\mathbf{s}))^* \underset{n\to\infty}{\xrightarrow{(w)}} (\sigma/2)((1 - s_1)\nu_2(\mathrm{d}\mathbf{s}))^*\,. \tag{24}$$

Now, for any non-negative measure μ on $\mathcal{S}^\downarrow$ and any non-negative continuous function f on $\mathcal{S}_1$, one can check that

$$((1 - s_1)\mu(\mathrm{d}\mathbf{s}))^*(f) = \int_{\mathcal{S}_1} \mu^*(\mathrm{d}\mathbf{x})(1 - \max \mathbf{x})f(\mathbf{x})\,, \tag{25}$$

where $\max \mathbf{x} = \max_{i\geq 1} x_i$. Applying (25) to $\mu = \overline{q}_n$ and ν_2, we conclude that (24) is a consequence of the following statement.

Proposition 15. *Let $f : \mathcal{S}_1 \to \mathbb{R}$ be a continuous function and let $g(\mathbf{x}) = (1 - \max \mathbf{x})f(\mathbf{x})$. Then*

$$\sqrt{n}\overline{q}^*_n(g) \underset{n\to\infty}{\longrightarrow} \frac{\sigma}{\sqrt{2\pi}} \int_0^1 \frac{\mathrm{d}x}{x^{1/2}(1-x)^{3/2}} g(x, 1-x, 0, \ldots) . \tag{26}$$

In summary, Theorem 3 in **Case 1.** is a consequence of this statement and Theorem 2.

Proposition 15 will be proved in a couple of steps. A difficulty that we will have to be careful about is that $\mathbf{x} \mapsto \max \mathbf{x}$ is not continuous on $\mathcal{S}_1$. Fix f, as in the statement. Note that $0 \le 1 - \max \mathbf{x} \le 1 - x_1$ for every $\mathbf{x} \in \mathcal{S}_1$, so that $g(\mathbf{x}) \le C(1 - x_1)$ for every $\mathbf{x} \in \mathcal{S}_1$ for some finite $C > 0$, a fact that will be useful.

First, note that combining (ii) in Proposition 14 with a size-biased ordering, it holds that

$$\overline{q}^*_n(g) = \sum_{p\ge1} q_n(p(\lambda) = p)\mathbb{E}\left[g\left(\frac{(X^*_1, \ldots, X^*_p, 0, \ldots)}{n} \right) \,\Big|\, \tau_p = n \right] . \tag{27}$$

Lemma 17. *For every $\varepsilon > 0$,*

$$\sqrt{n}q_n(p(\lambda) > \varepsilon\sqrt{n}) \underset{n\to\infty}{\longrightarrow} 0 .$$

Proof. From (23), the local limit theorem in the finite-variance case

$$\sup_{p\in\mathbb{Z}} \left| \sqrt{n}\mathbb{P}(S_n = -p) - \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{p^2}{2n\sigma^2} \right) \right| \underset{n\to\infty}{\longrightarrow} 0 , \tag{28}$$

shows that $q_n(p(\lambda) = p) \le C\hat{\xi}(p)$ for every n, p. Now

$$\sum_{k\ge0} \hat{\xi}((k, \infty)) < \infty ,$$

because $\hat{\xi}$ has finite mean. Since $\hat{\xi}((k, \infty))$ is non-increasing, this entails that $\hat{\xi}((k, \infty)) = o(k^{-1})$. Hence the result. $\square$

Lemma 18. *One has*

$$\lim_{\eta\downarrow0} \limsup_{n\to\infty} \sqrt{n}\overline{q}^*_n(|g|\mathbf{1}_{\{x_1>1-\eta\}}) = 0 , \qquad \textit{and} \qquad \lim_{n\to\infty} \sqrt{n}\overline{q}^*_n(\mathbf{1}_{\{x_1<n^{-7/8}\}}) = 0 .$$

Proof. Let $\varepsilon, \eta > 0$. Since $|f(\mathbf{x})| \le (1 - x_1)$, we get using (27), that $\sqrt{n}\overline{q}^*_n(|f|\mathbf{1}_{\{x_1>1-\eta\}})$ is bounded from above by

$$n^{1/2} \sum_{1\le p\le\varepsilon n^{1/2}} q_n(p(\lambda) = p) \sum_{m_1\ge(1-\eta)n} \left(1 - \frac{m_1}{n}\right) \frac{pm_1}{n} \frac{\mathbb{P}(X_1 = m_1)\mathbb{P}(\tau_{p-1} = n - m_1)}{\mathbb{P}(\tau_p = n)} + o(1) ,$$

because

$$\mathbb{P}(X^*_1 = m \,|\, X_1 + \ldots + X_p = n) = \frac{pm}{n} \frac{\mathbb{P}(X_1 = m)\mathbb{P}(X_2 + \ldots + X_p = n - m)}{\mathbb{P}(X_1 + \ldots + X_p = n)} .$$

The $o(1)$ term accounts for the fact that we restricted the sum to $1 \le p \le \varepsilon n^{1/2}$, which costs at most $o(n^{-1/2})$ by Lemma 17. We now use the cyclic lemma again, entailing that $\mathbb{P}(\tau_p = n) = (p/n)\mathbb{P}(S_n = -p)$. Using this for $p = 1$, and with the help of (28), we obtain that $\mathbb{P}(\tau_1 = n) \sim (\sigma\sqrt{2\pi})^{-1}n^{-3/2}$ as $n \to \infty$. This yields the upper-bound

$$\sqrt{n}\overline{q}^*_n(|g|\mathbf{1}_{\{x_1>1-\eta\}}) \le C \sum_{1\le p\le\varepsilon n^{1/2}} p^2\xi(p)\frac{1}{n} \sum_{(1-\eta)n\le m_1\le n} \frac{\sqrt{n}}{\sqrt{m_1}} \frac{\mathbb{P}(S_{n-m_1} = -p + 1)}{\mathbb{P}(S_n = -p)} + o(1) ,$$

and since $1 \le p \le \varepsilon n^{1/2}$, (28) implies that $\sqrt{n}\mathbb{P}(S_n = -p)$ and $\sqrt{n-m_1}\mathbb{P}(S_{n-m_1} = -p+1)$ are respectively bounded from below and above, by constants that are independent on n, m_1, p (but might depend on ε). Consequently, the bound is

$$C_\varepsilon \sum_{p\ge 1} p^2\xi(p)\frac{1}{n}\sum_{(1-\eta)n<m_1\le n}\frac{1}{\sqrt{\frac{m_1}{n}\left(1-\frac{m_1}{n}\right)}} + o(1)\,,$$

and this converges to $C_\varepsilon \int_{1-\eta}^{1}(x(1-x))^{-1/2}\mathrm{d}x$. In turn, this goes to 0 as $\eta \to 0$, for fixed ε. The second limit is obtained in a similar way, writing $\sqrt{n}\overline{q}_n^*(\mathbf{1}_{\{x_1<n^{-7/8}\}})$ as

$$\begin{aligned}
& n^{1/2}\sum_{1\le p\le \varepsilon n^{1/2}} q_n(p(\lambda)=p)\sum_{1\le m_1\le n^{1/8}}\frac{pm_1}{n}\frac{\mathbb{P}(X_1=m_1)\mathbb{P}(\tau_{p-1}=n-m_1)}{\mathbb{P}(\tau_p=n)} + o(1)\\
\le\ & n^{-3/8}\sum_{1\le p\le \varepsilon n^{1/2}} p(p-1)\xi(p)\sum_{1\le m_1\le n^{1/8}}\frac{\mathbb{P}(S_{n-m_1}=-p+1)}{\mathbb{P}(S_{n+1}=-1)} + o(1)\\
\le\ & C_\varepsilon n^{-1/4}\sum_{p\ge 1} p(p-1)\xi(p) + o(1)\,,
\end{aligned}$$

for some constant C_ε, where we used the local limit theorem at the last step. □

Lemma 19. *For every $\eta > 0$, it holds that*

$$\lim_{n\to\infty}\sqrt{n}\overline{q}_n^*(\mathbf{1}_{\{x_1+x_2<1-\eta\}}) = 0\,.$$

Proof. Fix $\varepsilon > 0$. An upper-bound for the quantity appearing in the statement is given by (up to an additional $o(1)$ quantity depending on ε)

$$\sqrt{n}\sum_{1\le p\le \varepsilon n^{1/2}} p\xi(p)\sum_{m_1+m_2\le(1-\eta)n}\frac{pm_1}{n}\frac{(p-1)m_2}{n-m_1}\frac{\mathbb{P}(X_1=m_1)\mathbb{P}(X_2=m_2)\mathbb{P}(\tau_{p-2}=n-m_1-m_2)}{\mathbb{P}(\tau_p=n)}\,.$$

If $m_1+m_2 < n(1-\eta)$ then $n-m_1-m_2 \ge \eta n$. In this case, we obtain, using the cyclic lemma and (28),

$$\frac{\mathbb{P}(\tau_{p-2}=n-m_1-m_2)}{\mathbb{P}(\tau_p=n)} = \frac{\frac{p-2}{n-m_1-m_2}\mathbb{P}(S_{n-m_1-m_2}=-p+2)}{\frac{p}{n}\mathbb{P}(S_n=-p)} \le \frac{C}{\eta^{3/2}}\,.$$

Note that the constant C here does not depend on p, ε. Consequently, we obtain the bound

$$\begin{aligned}
\sqrt{n}\overline{q}_n^*(\mathbf{1}_{\{x_1+x_2<(1-\eta)n\}}) &\le \frac{C}{\eta^{3/2}\sqrt{n}}\sum_{1\le p\le\varepsilon n^{1/2}} p^3\xi(p)\frac{1}{n^2}\sum_{m_1+m_2<(1-\eta)n}\sqrt{\frac{n}{m_1}\frac{n}{m_2}} + o(1)\\
&\le \frac{C\varepsilon}{\eta^{3/2}}\sum_{p\ge 1}p^2\xi(p)\int_{x_1+x_2\le 1}\frac{\mathrm{d}x_1\mathrm{d}x_2}{\sqrt{x_1x_2}} + o(1)\,,
\end{aligned}$$

where C is still independent of p, ε. The first term on the right-hand side does not depend on n anymore and goes to 0 as $\varepsilon \to 0$, entailing the result. □

Lemma 20. *There exists a function $\beta_\eta = o(\eta)$ as $\eta \downarrow 0$, so that*

$$\begin{aligned}
\lim_{\eta\downarrow 0}\liminf_{n\to\infty}\sqrt{n}\overline{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta, x_1+x_2>1-\beta_\eta\}}) &= \lim_{\eta\downarrow 0}\limsup_{n\to\infty}\sqrt{n}\overline{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta, x_1+x_2>1-\beta_\eta\}})\\
&= \frac{\sigma}{\sqrt{2\pi}}\int_0^1\frac{g((x,1-x,0,\ldots))}{x^{1/2}(1-x)^{3/2}}\mathrm{d}x\,,
\end{aligned}$$

Proof. The proof is similar to the previous ones, but technically more tedious, so we will only sketch the details. Fix $\eta > 0$, and consider $\eta' \in (0, \eta)$ and $\varepsilon > 0$. Then, by decomposing with respect to the events $\{p(\lambda) > \varepsilon\sqrt{n}\}$ and $\{\mathbf{x} : x_1 \leq n^{-7/8}\}$, we obtain, using Lemma 17 and the second limit of Lemma 18,

$$\begin{aligned}
\sqrt{n}\overline{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta, x_1+x_2>1-\eta'\}}) &= o(1) + \sqrt{n} \sum_{1\leq p\leq \varepsilon n^{1/2}} q_n(p(\lambda) = p) \\
\times & \sum_{\substack{n^{1/8}\leq m_1\leq (1-\eta)n \\ (1-\eta')n\leq m_1+m_2\leq n}} \mathbb{E}[g((m_1, m_2, X_3^*, \ldots, X_p^*)/n) \,|\, \tau_p = n, X_1^* = m_1, X_2^* = m_2] \\
\times & \frac{pm_1}{n} \frac{(p-1)m_2/n}{1-m_1/n} \mathbb{P}(X_1 = m_1)\mathbb{P}(X_2 = m_2) \frac{\mathbb{P}(\tau_{p-2} = n - m_1 - m_2)}{\mathbb{P}(\tau_p = n)}\,.
\end{aligned}$$

We now give a lower bound of the $\liminf$ of this as $n \to \infty$. Showing that the same quantity is an upper-bound of the $\limsup$ being similar and easier.

Note that if $x_1 + x_2 \geq 1 - \eta'$ and $x_1 \leq 1 - \eta$, we have that $(1 - x_1 - x_2)/(1 - x_1) \leq \eta'/\eta$ and then $x_2/(1 - x_1) \geq 1 - \eta'/\eta$. Next, by the local limit theorem, we can always choose η' small enough so that $\mathbb{P}(X_1 = m_2)/\mathbb{P}(X_1 = n - m_1) \geq 1 - \eta$ for every n large enough.

Also, still by the local limit theorem, we can choose ε small enough so that for every $1 \leq p \leq \varepsilon n^{1/2}$ and every n large, we have

$$q_n(p(\lambda) = p)/\hat{\xi}(p) \geq (1 - \eta) \qquad \text{and} \qquad p^{-1}n^{3/2}\mathbb{P}(\tau_p = n) \geq (1 - \eta)\sigma\sqrt{2\pi}\,.$$

A third use of the local limit theorem entails that for every n large, we have

$$m_1^{3/2}\mathbb{P}(X_1 = m_1) \wedge m_2^{3/2}\mathbb{P}(X_2 = m_2) \geq (1 - \eta)\sigma\sqrt{2\pi}\,,$$

for every n large and $m_1 \geq n^{1/8}$, $m_2 \geq (\eta - \eta')n$.

Finally, we use the fact that f is uniformly continuous on $\mathcal{S}_1$, while $\max \mathbf{x} = x_1 \vee x_2$ on the set $\{\mathbf{x} \in \mathcal{S}_1 : x_1 + x_2 > 3/4\}$. Consequently, the function $g(\mathbf{x}) = (1 - \max \mathbf{x})f(\mathbf{x})$ is uniformly continuous on the latter set. Therefore, we can choose $\eta' < 1/4$ small enough so that

$$|g((m_1, m_2, m_3, \ldots)/n) - g((m_1, n - m_1, 0, \ldots)/n)| \leq \eta$$

for every $(m_1, m_2, \ldots)$ with sum n, such that $m_1 + m_2 \geq (1 - \eta')n$. Putting things together, for every $\eta > 0$, we can choose $\eta' =: \beta_\eta, \varepsilon$ small so that for every n large enough, $\sqrt{n}\overline{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta,x_1+x_2>1-\eta'\}})$ is greater than or equal to

$$\begin{aligned}
(1 - \eta)^5(1 - \eta'/\eta) & \sum_{1\leq p\leq \varepsilon n^{1/2}} (p-1)\hat{\xi}(p)\frac{1}{n} \sum_{n^{1/8}\leq m_1\leq (1-\eta)n} (g((m_1, n - m_1, 0, \ldots)/n) - \eta)\frac{m_1}{n} \\
\times & \frac{1}{\sigma\sqrt{2\pi}((m_1/n)(1 - m_1/n))^{3/2}} \sum_{(1-\eta')n-m_1\leq m_2\leq n-m_1} \mathbb{P}(\tau_{p-2} = n - m_1 - m_2).
\end{aligned}$$

Finally, the last sum is $\sum_{m=0}^{\eta' n} \mathbb{P}(\tau_{p-2} = m)$, and this can be made arbitrarily close to 1, uniformly in $1 \leq p \leq \varepsilon n^{1/2}$, as soon as n is large enough, by our usual use of the local limit theorem. Taking the $\liminf$ in n and using a convergence of Riemann sums, yields

$$\begin{aligned}
&\liminf_{n\to\infty} \sqrt{n}\overline{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta,x_1+x_2>1-\eta'\}}) \\
&\quad\geq (1 - \eta)^5(1 - \eta'/\eta) \sum_{p\geq 1} (p-1)\hat{\xi}(p) \int_0^{1-\eta} \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}x^{1/2}(1 - x)^{3/2}} (g(x, 1 - x, 0, \ldots) - \eta).
\end{aligned}$$

One concludes using the fact that $\sum_{p\geq 1}(p-1)\hat{\xi}(p)=\sigma^2$. $\square$

We can now finish the proof of Proposition 15. Simply write

$$|\overline{q}_n^*(g)-\overline{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta,x_1+x_2>1-\eta'\}})|\leq \overline{q}_n^*(|g|\mathbf{1}_{\{x_1\geq 1-\eta\}}))+\overline{q}_n^*(|g|\mathbf{1}_{\{x_1+x_2\leq 1-\eta'\}})\,.$$

Now fix $\varepsilon>0$, and using Lemmas 18 and 20, choose η,η' in such a way that $\sqrt{n}\overline{q}_n^*(|g|\mathbf{1}_{\{x_1\geq 1-\eta\}})\leq\varepsilon/2$ and

$$\left|\sqrt{n}\overline{q}_n^*(g\mathbf{1}_{\{x_1<1-\eta,x_1+x_2>1-\eta'\}})-\frac{\sigma}{\sqrt{2\pi}}\int_0^1\frac{g((x,1-x,0,\ldots))}{x^{1/2}(1-x)^{3/2}}\mathrm{d}x\right|\leq\varepsilon/2$$

for every n large. For this choice of η,η', we then have for every n large enough,

$$\left|\overline{q}_n^*(g)-\frac{\sigma}{\sqrt{2\pi}}\int_0^1\frac{g((x,1-x,0,\ldots))}{x^{1/2}(1-x)^{3/2}}\mathrm{d}x\right|\leq\varepsilon+\overline{q}_n^*(|g|\mathbf{1}_{\{x_1+x_2\leq 1-\eta'\}})\,,$$

and the upper-bound converges to ε as $n\to\infty$ by Lemma 19. Since ε was arbitrary, this proves Proposition 15, hence implying Theorem 3 in **Case 1**.

5.2 Stable case

Assume that $\xi(p)\sim cp^{-\alpha-1}$ for some $\alpha\in(1,2)$ and $c>0$. Theorem 3 in **Case 2.** will follow if we can show that hypothesis **(H)** holds for $\gamma=1-1/\alpha$, $\ell\equiv(\alpha(\alpha-1)/(c\Gamma(2-\alpha)))^{1/\alpha}$, and the dislocation measure ν_α. A similar reasoning as in the beginning of the previous section shows that it suffices to prove the following statement.

Proposition 16. *If $f:\mathcal{S}_1\to\mathbb{R}$ is a continuous function and $g(\mathbf{x})=(1-\max\mathbf{x})f(\mathbf{x})$, then*

$$n^{1-1/\alpha}\overline{q}_n^*(g)\underset{n\to\infty}{\longrightarrow}\left(c\frac{\Gamma(2-\alpha)}{\alpha(\alpha-1)}\right)^{1/\alpha}\nu_\alpha^*(g)\,.$$

One will note that the function g of the statement is continuous ν_α^*-a.e., since $\mathbf{x}\mapsto\max\mathbf{x}$ is continuous at every point $\mathbf{x}$ with sum 1. Now,

$$\begin{aligned}\overline{q}_n^*(g) &= \sum_{p\geq 1}q_n(p(\lambda)=p)\mathbb{E}\left[g\left(\frac{(X_1^*,\ldots,X_p^*,0,\ldots)}{n}\right)\,\Big|\,\tau_p=n\right]\\ &= n^{1/\alpha}\int_0^\infty\mathrm{d}x\,q_n(p(\lambda)=\lceil n^{1/\alpha}x\rceil)\,\mathbb{E}\left[g\left(\frac{(X_1^*,\ldots,X_{\lceil n^{1/\alpha}x\rceil}^*,0,\ldots)}{n}\right)\,\Big|\,\tau_{\lceil n^{1/\alpha}x\rceil}=n\right]\end{aligned}\tag{29}$$

Recall the notation around (23). The random walk S_n is now such that $(S_{\lfloor nt\rfloor}/n^{1/\alpha},t\geq 0)$ converges in distribution in the Skorokhod space to a spectrally positive stable Lévy process $(Y_t,t\geq 0)$ with index α and Lévy measure $c\mathrm{d}x/x^{1+\alpha}\mathbf{1}_{\{x>0\}}$. Its Laplace transform is given by $\mathbb{E}[\exp(-\lambda Y_t)]=\exp(tc'\lambda^\alpha)$, where $c'=c\frac{\Gamma(2-\alpha)}{\alpha(\alpha-1)}$. The Gnedenko-Kolmogorov local limit theorem also yields

$$n^{1/\alpha}\mathbb{P}(S_n=k)=p_1(k/n^{1/\alpha})+\varepsilon(n,k)$$

where $\sup_k|\varepsilon(n,k)|\to 0$ as $n\to\infty$, and p_t is the density of Y_t. This, together with (23) and our hypothesis on the asymptotic behavior of ξ, entail that

$$q_n(p(\lambda)=\lceil n^{1/\alpha}x\rceil)\sim cn^{-1}x^{-\alpha}\frac{p_1(-x)}{p_1(0)}\,.$$

Let us now focus on the random variables $X_1,X_2,\ldots$ and $\tau_p=X_1+\ldots+X_p$. We have $\mathbb{P}(X_1=n)=n^{-1}\mathbb{P}(S_n=-1)\sim n^{-1-1/\alpha}p_1(0)$, which gives that X_1 is in the domain of attraction of a stable

random variable with index $1/\alpha$. More specifically, one has that $(\tau_{\lfloor nx\rfloor}/n^\alpha, x\geq 0)$ converges in the Skorokhod space to a stable subordinator $(T_y, y\geq 0)$ with index $1/\alpha$, and Lévy measure

$$p_1(0)\frac{\mathrm{d}x}{x^{1+1/\alpha}}\mathbf{1}_{\{x>0\}}\,. \tag{30}$$

Its Laplace transform is given by

$$\mathbb{E}[\exp(-\lambda T_x)] = \exp(-xp_1(0)\alpha\Gamma(1-1/\alpha)\lambda^{1/\alpha})\,.$$

On the other hand, T_x has same distribution as the first hitting time of $-x$ by $(Y_t, t\geq 0)$ (because a similar statement is true of τ_p and S_n), which identifies the Laplace exponent of T_1 as $(\lambda/c')^{1/\alpha}$, see [7, Chapter VII]. This yields

$$p_1(0) = \frac{1}{\alpha\Gamma(1-1/\alpha)(c')^{1/\alpha}} = \frac{1}{\alpha\Gamma(1-1/\alpha)}\Big(\frac{\alpha(\alpha-1))}{c\Gamma(2-\alpha)}\Big)^{1/\alpha}. \tag{31}$$

Let Q_y be the probability density function of T_y. The cyclic lemma [7, Corollary VII.3] gives $tQ_x(t) = xp_t(-x)$, while the Gnedenko-Kolmogorov local limit theorem states that

$$p^\alpha \mathbb{P}(\tau_p = n) = Q_1(n/p^\alpha) + \varepsilon'(p,n)\,,$$

where $\sup_n|\varepsilon'(p,n)|\to 0$ as $p\to\infty$.

Lemma 21. *The sequence $(X_1^*,\ldots,X^*_{\lceil n^{1/\alpha}x\rceil})/n$ conditioned on $\tau_{\lceil n^{1/\alpha}x\rceil} = n$ converges in distribution to a random sequence $(\Delta_1^*,\Delta_2^*,\ldots)$, defined inductively by*

$$\mathbb{P}\Big(\Delta^*_{i+1}\in \mathrm{d}z\,\Big|\,\Delta^*_1,\ldots,\Delta^*_i,\sum_{j=1}^i\Delta^*_j = y\Big) = \frac{p_1(0)x}{z^{1/\alpha}}\frac{Q_x(1-y-z)}{Q_x(1-y)}\mathrm{d}z\,, \qquad 0\leq z\leq 1-y\,.$$

Proof. The case $i=1$ is obtained by using the local limit theorem in

$$n\mathbb{P}(X_1^* = \lfloor nz\rfloor\,|\,\tau_{\lceil n^{1/\alpha}x\rceil} = n) = \lceil n^{1/\alpha}x\rceil\,\lfloor nz\rfloor\mathbb{P}(X_1 = \lfloor nz\rfloor)\frac{\mathbb{P}(\tau_{\lceil n^{1/\alpha}x\rceil-1} = n-\lfloor nz\rfloor)}{\mathbb{P}(\tau_{\lceil n^{1/\alpha}x\rceil} = n)}\,.$$

One then reasons inductively, in an elementary way. Details are left to the reader. □

The limiting sequence $(\Delta^*_i, i\geq 1)$ has same distribution as the sequence of jumps of the subordinator $(T_y, 0\leq y\leq x)$, conditionally given $T_x = 1$, and arranged in size-biased order, see [32, Chapter 4] or [9]. We will denote by $\Delta T^*_{[0,x]}$ this randomly ordered sequence of jumps. Hence, provided we have the right to apply dominated convergence in (29), we obtain, using $xp_1(-x) = Q_x(1)$,

$$n^{1-1/\alpha}\overline{q}^*_n(g) \underset{n\to\infty}{\longrightarrow} \frac{c}{p_1(0)}\int_0^\infty \frac{\mathrm{d}x}{x^{\alpha+1}}Q_x(1)\mathbb{E}[g(\Delta T^*_{[0,x]})\,|\,T_x = 1]\,. \tag{32}$$

Using scaling for the subordinator $(T_y, y\geq 0)$, the previous integral can be rewritten as

$$\frac{c}{p_1(0)}\int_0^\infty \frac{\mathrm{d}x}{x^{2\alpha+1}}Q_1(x^{-\alpha})\mathbb{E}[g(x^\alpha\Delta T^*_{[0,1]})\,|\,T_1 = x^{-\alpha}]\,,$$

and changing variables $u = x^{-\alpha}$ shows that this is equal to

$$\frac{c}{\alpha p_1(0)}\int_0^\infty Q_1(u)\mathrm{d}u\,\mathbb{E}[ug(\Delta T^*_{[0,1]}/u)\,|\,T_1 = u] = \frac{c}{\alpha p_1(0)}\mathbb{E}[T_1 g(\Delta T^*_{[0,1]}/T_1)]\,.$$

Finally, the sequence $\Delta T_{[0,1]}$ of jumps of T before time 1 is the sequence of atoms of a Poisson measure with intensity given by (30). Using (31), it thus has same distribution as $\alpha(\alpha-1)c^{-1}\Gamma(2-\alpha)^{-1}(\Delta_1,\Delta_2,\ldots)$, as defined in Section 2.1. Using the notations therein and (31), we get after rearrangements

$$\begin{aligned}\frac{c}{\alpha p_1(0)}\mathbb{E}[T_1 g(\Delta T^*_{[0,1]}/T_1)] &= \left(c\frac{\Gamma(2-\alpha)}{\alpha(\alpha-1)}\right)^{1/\alpha}\frac{\alpha^2\Gamma(2-1/\alpha)}{\Gamma(2-\alpha)}\mathbb{E}\left[Tg\left(\frac{\Delta_i^*}{T},i\geq 1\right)\right]\\ &= \left(c\frac{\Gamma(2-\alpha)}{\alpha(\alpha-1)}\right)^{1/\alpha}\nu^*_\alpha(g)\,,\end{aligned}$$

as wanted. It remains to justify that the convergence (32) is indeed dominated. To this end, using (29) and the fact that $q_n(p(\lambda)=\lceil n^{1/\alpha}x\rceil)\leq C\lceil n^{1/\alpha}x\rceil^{-\alpha}$, it suffices to show that the expectation term in this equation is bounded by $C\lceil n^{1/\alpha}x\rceil/n^{1/\alpha}$ for some C independent of n, and for $x\in[0,1]$. In turn, since $g(\mathbf{x})\leq C(1-x_1)$ for some finite $C>0$, it suffices to substitute this upper-bound to g. Now, we have $\mathbb{P}(X_1=m)\leq Cm^{-1-1/\alpha}$ for every m, so that

$$\begin{aligned}&\mathbb{E}\left[\left(1-\frac{X_1^*}{n}\right)\,\Big|\,\tau_{\lceil n^{1/\alpha}x\rceil}=n\right]\\ &= \sum_{m=1}^{n}\left(1-\frac{m}{n}\right)\lceil n^{1/\alpha}x\rceil\frac{m}{n}\mathbb{P}(X_1=m)\frac{\mathbb{P}(\tau_{\lceil n^{1/\alpha}x\rceil-1}=n-m)}{\mathbb{P}(\tau_{\lceil n^{1/\alpha}x\rceil}=n)}\\ &\leq \sum_{m=1}^{n}\left(1-\frac{m}{n}\right)\lceil n^{1/\alpha}x\rceil\frac{m}{n}\mathbb{P}(X_1=m)\frac{\frac{\lceil n^{1/\alpha}x\rceil-1}{n-m}\mathbb{P}(S_{n-m}=-\lceil n^{1/\alpha}x\rceil+1)}{\frac{\lceil n^{1/\alpha}x\rceil}{n}\mathbb{P}(S_n=-\lceil n^{1/\alpha}x\rceil)}\\ &\leq C\frac{\lceil n^{1/\alpha}x\rceil}{n^{1/\alpha}}\frac{1}{n}\sum_{m=1}^{n}\frac{1}{\left(\frac{m}{n}\right)^{1/\alpha}\left(1-\frac{m}{n}\right)^{1/\alpha}}\\ &\leq C\frac{\lceil n^{1/\alpha}x\rceil}{n^{1/\alpha}}\,,\end{aligned}$$

where we have used that $(n-m)^{1/\alpha}\mathbb{P}(S_{n-m}=-\lceil n^{1/\alpha}x\rceil+1)$ is uniformly bounded (in n,m,x) and that $n^{1/\alpha}\mathbb{P}(S_n=-\lceil n^{1/\alpha}x\rceil)$ is uniformly bounded away from 0 for $x\in[0,1]$. This is the wanted bound, concluding the proof of Proposition 16, hence of Theorem 3.

6 Scaling limits of uniform unordered trees

In this section, we fix once and for all an integer $m\in\{2,\ldots,\infty\}$ and consider trees in which every vertex has at most m children. We use the notations of Section 2.2 and let T_n be uniformly distributed in $\mathsf{T}_n^{(m)}$, for $n\geq 1$.

The first difficulty we have to overcome is that the sequence $(T_n,n\geq 1)$ *is not* Markov branching as defined in Section 1.2.2. We will therefore start in Section 6.1 by coupling this sequence with a family of Markov branching trees that are asymptotically close to $T_n,n\geq 1$, and then check in Section 6.2 that the coupled trees satisfy **(H)**.

Let us fix some notation. For $\mathsf{t}\in\mathsf{T}_n^{(m)}$, we can write $\mathsf{t}=\langle\mathsf{t}^{(1)},\ldots,\mathsf{t}^{(k)}\rangle$ with $\sum_{i=1}^k\#\mathsf{t}^{(i)}=n-1$, and we let $\lambda(\mathsf{t})\in\mathcal{P}_{n-1}$ be the partition obtained by arranging in decreasing order the sequence $(\#\mathsf{t}^{(1)},\ldots,\#\mathsf{t}^{(k)})$ (of course, this does not depend on the labeling of the trees $\mathsf{t}^{(1)},\ldots,\mathsf{t}^{(k)}$). Let $\mathsf{F}_j(k)$ be the set of of multisets[3] with k elements in $\mathsf{T}_j^{(m)}$. By convention, we set $\mathsf{F}_j(0)=\{\varnothing\}$.

[3]Recall that a multiset with k elements in some set A is an element of the quotient set $A^k/\mathfrak{S}_k$, where $\mathfrak{S}_k$ acts in the natural way by permutation of components.

Then, for $\lambda \in \mathcal{P}_{n-1}$ with $p(\lambda) \leq m$, we have a bijection

$$\{\mathsf{t} \in \mathsf{T}_n^{(m)} : \lambda(\mathsf{t}) = \lambda\} \equiv \prod_{j=1}^{n-1} \mathsf{F}_j(m_j(\lambda)) \,, \tag{33}$$

obtained by grouping the subtrees of t born from the root with size j into a multiset, denoted by $\mathsf{f}_j(\mathsf{t})$, of $m_j(\lambda)$ trees. From this, we deduce that $\mathsf{f}_j(T_n), 1 \leq j \leq n-1$ are independent uniform random elements in $\mathsf{F}_j(m_j(\lambda))$ conditionally given $\lambda(T_n)$. However, the uniform random element in $\mathsf{F}_j(k)$ has a different distribution from the multiset induced by k i.i.d. uniform element in $\mathsf{T}_j^{(m)}$, as soon as $k \geq 2$. This is what prevents T_n from enjoying the Markov branching property, i.e. from having law Q_n^q, where for $n \geq 1$, q_n is the law of $\lambda(T_{n+1})$.

Letting $\mathbf{F}_j(k) = \#\mathsf{F}_j(k)$, the previous bijection yields

$$\mathbf{S}_n^{(\lambda)} := \#\{\mathsf{t} \in \mathsf{T}_n^{(m)} : \lambda(\mathsf{t}) = \lambda\} = \prod_{j=1}^{n-1} \mathbf{F}_j(m_j(\lambda)) \,.$$

When $p(\lambda) \geq m$, we set $\mathbf{S}_n^{(\lambda)} = 0$. Of course, letting $\mathbf{T}_n^{(m)} = \#\mathsf{T}_n^{(m)}$, we also have

$$\mathbf{T}_n^{(m)} = \sum_{\lambda \in \mathcal{P}_{n-1}} \mathbf{S}_n^{(\lambda)} \,.$$

Using the obvious fact that $\mathbf{F}_j(k) \leq \mathbf{T}_j^{(m)} \mathbf{F}_j(k-1)$, we obtain the rough but useful bound

$$\mathbf{S}_n^{(\lambda)} \leq \mathbf{T}_{\lambda_1}^{(m)} \mathbf{S}_{n-\lambda_1}^{(\lambda_2,\lambda_3,\ldots,\lambda_{p(\lambda)})} . \tag{34}$$

We recall the key result (5) of Otter [31], which is used throughout the proofs below:

$$\mathbf{T}_n^{(m)} \underset{n\to\infty}{\sim} \kappa_m \frac{\rho^n}{n^{3/2}} .$$

Setting $\mathbf{T}_0^{(m)} = 1$ by convention, we obtain that for $\rho = \rho_m > 1$, and two constants $K \geq 1 \geq \mathrm{k} > 0$,

$$\mathbf{T}_n^{(m)} \leq K \frac{\rho^n}{n^{3/2}} \,, \quad n \geq 0 \qquad \mathbf{T}_n^{(m)} \geq \mathrm{k} \frac{\rho^n}{n^{3/2}} , \quad n \geq 1 \,. \tag{35}$$

Note that we also have $\mathbf{T}_n^{(m)} \leq K \rho^n$ for all $n \geq 0$. Last, we let $\kappa = \kappa_m$.

6.1 Coupling

Let μ_n be the uniform probability distribution over $\mathsf{T}_n^{(m)}$, and let $q_n = \lambda_* \mu_{n+1}$ be the law of the partition of n induced by the subtrees born from the root of a μ_{n+1}-distributed tree. For every $n \geq 1$, we want to construct a pair of random variables (T_n, T'_n) on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that

- T_n has law μ_n
- T'_n has law Q_n^q
- for every $\varepsilon > 0$, $\lim_{n\to\infty} \mathbb{E}\left[d_{\mathrm{GHP}}(n^{-\varepsilon} T_n, n^{-\varepsilon} T'_n)\right] = 0$.

Recall that if T_n has distribution μ_n, and conditionally on $\lambda(T_n) = \lambda$, then $\mathsf{f}_j(T_n), 1 \leq j \leq n-1$ are independent, respectively uniform in $\mathsf{F}_j(m_j(\lambda))$. We are going to need the following fact.

Lemma 22. *For every $j, k \geq 1$, let F_j be uniform in $\mathsf{F}_j(k)$ and $\overline{F}_j$ be the multiset induced by an i.i.d. sequence of k random variables with law μ_j. Let A_j be the set of elements in $\mathsf{F}_j(k)$ with components that are pairwise distinct. Then*
(i) *One has $\mathbb{P}(F_j \in A_j) \leq \mathbb{P}(\overline{F}_j \in A_j)$.*
(ii) *The conditional distributions of F_j and $\overline{F}_j$ given A_j are equal.*

Proof. For a finite set A, the number of multisets with k elements is $\#(A^k/\mathfrak{S}_k) \geq \#A^k/k!$. Then

$$\begin{aligned}\mathbb{P}(F_j \in A_j) &= \frac{\#\mathsf{T}_j^{(m)}(\#\mathsf{T}_j^{(m)}-1)\dots(\#\mathsf{T}_j^{(m)}-k+1)}{k!\#\mathsf{F}_j(k)} \\ &\leq \frac{\#\mathsf{T}_j^{(m)}(\#\mathsf{T}_j^{(m)}-1)\dots(\#\mathsf{T}_j^{(m)}-k+1)}{(\#\mathsf{T}_j^{(m)})^k} \\ &= \mathbb{P}(\overline{F}_j \in A_j)\,.\end{aligned}$$

This gives (i). Property (ii) is also obtained by counting: on the event A_j. The probability that F_j equals some given (multi)set $S \in \mathsf{F}_j(k)$ with all distinct elements is $\#\mathsf{F}_j(k)^{-1}$, while the probability that F_j' equals the same set S is $k!(\#\mathsf{T}_j^{(m)})^{-k}$. Dividing by $\mathbb{P}(F_j \in A_j)$ and $\mathbb{P}(F_j' \in A_j)$ respectively gives the same result. $\square$

The previous statement allows to construct a coupling between F_j and $\overline{F}_j$, in the following way. Let $\mathsf{f} \in \mathsf{F}_j(k)$. Consider three independent random variables $\mathsf{f}'', \mathsf{f}''', B$, such that the law of f'' is the law of $\overline{F}_j$ conditionally given A_j, the law of f''' is the law of $\overline{F}_j$ conditionally given A_j^c and B is an independent Bernoulli random variable with $\mathbb{P}(B=1) = \mathbb{P}(\overline{F}_j \in A_j^c)/\mathbb{P}(F_j \in A_j^c)$, which is indeed in $[0,1]$ by (i) in Lemma 22. Set

$$\mathsf{f}' = \begin{cases} \mathsf{f} & \text{if } \mathsf{f} \in A_j \\ \mathsf{f}'' & \text{if } \mathsf{f} \notin A_j \text{ and } B = 0 \\ \mathsf{f}''' & \text{if } \mathsf{f} \notin A_j \text{ and } B = 1. \end{cases}$$

We let $K_j(\mathsf{f}, \cdot)$ be the law of the multiset f' thus obtained, hence defining a Markov kernel on $\mathsf{F}_j(k)$. We say that the random variables F, F' are *naturally coupled* if (F, F') has law $\mu(\mathrm{d}\mathsf{f})K_j(\mathsf{f}, \mathrm{d}\mathsf{f}')$, where μ is the law of F on $\mathsf{F}_j(k)$. Using (ii) in Lemma 22, it is then easy to obtain the next result.

Lemma 23. *If F_j is uniform in $\mathsf{F}_j(k)$ and (F_j, F_j') are naturally coupled, then the law of F_j' is that of the multiset induced by k i.i.d. uniform elements in T_j.*

Next, we define a Markov kernel $K(\mathsf{t}, \cdot)$ on $\mathsf{T}^{(m)}$, in an inductive way. Let $K(\bullet, \{\bullet\}) = 1$. Assume that the measure $K(\mathsf{t}, \cdot)$ on $\mathsf{T}_{\#\mathsf{t}}^{(m)}$ has been defined for every $\mathsf{t} \in \mathsf{T}^{(m)}$ with $\#\mathsf{t} \leq n-1$. Take $\mathsf{t} \in \mathsf{T}_n^{(m)}$, and let $\lambda = \lambda(\mathsf{t}), p = p(\lambda)$. Let $\mathsf{f}_j(\mathsf{t}) \in \mathsf{F}_j(m_j(\lambda)), 1 \leq j \leq n-1$ be the multisets of trees born from the root of t, respectively with size j. Let $\mathsf{f}_j'(\mathsf{t})$ be independent random multisets, respectively with law $K_j(\mathsf{f}_j(\mathsf{t}), \cdot)$. We relabel the p elements of the multisets $\mathsf{f}_j'(\mathsf{t}), 1 \leq j \leq n-1$ as $\mathsf{t}_{(1)}, \dots, \mathsf{t}_{(p)}$, in non-increasing order of size, so that $\#\mathsf{t}_{(i)} = \lambda_i$ — if there is some j with $m_j(\lambda) \geq 2$, we arrange the trees with same size in exchangeable random order. All these trees are in $\mathsf{T}^{(m)}$ and have at most $n-1$ vertices. By the induction hypothesis, conditionally on this family, we can consider another family $\mathsf{t}_{(1)}', \dots, \mathsf{t}_{(p)}'$ of independent trees with respective laws $K(\mathsf{t}_{(i)}, \cdot)$. Let $K(\mathsf{t}, \cdot)$ be the law of the tree $\langle \mathsf{t}_{(i)}', 1 \leq i \leq p \rangle$. This procedure allows to define the Markov kernel $K(\mathsf{t}, \cdot)$ for every tree in $\mathsf{T}^{(m)}$.

We say that the random trees (T, T'), defined on a common probability space, are *naturally coupled* if the law of (T, T') is $\mu(\mathrm{d}\mathsf{t})K(\mathsf{t}, \mathrm{d}\mathsf{t}')$, where μ is the law of T. Is is easy to see that for every random variable T on $\mathsf{T}^{(m)}$ with law μ, then, possibly to the cost of enlarging the probability space supporting T, one can construct a random variable T' so that (T, T') is naturally coupled.

Proposition 17. *Let T_n have law μ_n and (T_n, T'_n) be naturally coupled. Endow these trees respectively with the measures μ_{T_n} and $\mu_{T'_n}$. Then,*
(i) *the tree T'_n has distribution Q^q_n, for every $n \geq 1$,*
(ii) *for all $a > 0$, the Gromov-Hausdorff-Prokhorov distance between aT_n and aT'_n is at most $2aj^*$ where j^* is the supremum integer $j \geq 1$ so that there exist two subtrees of T_n with size j, born from the same vertex and which are equal (with the convention $\sup \varnothing = 0$),*
(iii) *for all $\varepsilon > 0$, $\mathbb{E}\left[d_{\mathrm{GHP}}(n^{-\varepsilon}T_n, n^{-\varepsilon}T'_n)\right] \to 0$ as $n \to \infty$.*

Proof. We prove (i) by induction. For $n = 1$ the property is obvious. Assume the property holds for every index up to $n-1$, and condition on $\lambda(T_n) = \lambda$, which by definition has probability $q_{n-1}(\lambda)$. As noticed before Lemma 22, the multisets $F_j = \mathsf{f}_j(T_n), 1 \leq j \leq n-1$ are independent, respectively uniform in $\mathsf{F}_j(m_j(\lambda))$. Conditionally on $F_j, 1 \leq j \leq n-1$, let $F'_j, 1 \leq j \leq n-1$ be independent with respective laws $K_j(F_j, \cdot)$. By Lemma 23, we obtain that F'_j is the multiset induced by a sequence of $m_j(\lambda)$ i.i.d. random variables, with law μ_j. Consequently, if we relabel the elements of $F'_j, 1 \leq j \leq n-1$ as $T_{(1)}, \ldots, T_{(p)}$ in non-increasing order of size (and exchangeable random order for trees with same size), then we obtain that these trees are independent, respectively with distribution μ_{λ_j}. Since, by definition of K, the natural coupling (T_n, T'_n) is obtained by letting $T'_n = \langle T'_{(i)}, 1 \leq i \leq p\rangle$ where $(T_{(i)}, T'_{(i)})$ are naturally coupled, we readily obtain the Markov branching property, with branching laws $(q_n, n \geq 1)$.

For (ii), we again apply an induction argument. The statement is trivial for $n = 1$. Now, in the first step of the natural coupling, the action of the Markov kernel K_j on $\mathsf{f}_j(T_n)$ leaves it unchanged if $\mathsf{f}_j(T_n) \in A_j$, i.e. if there are no ties in the multiset $\mathsf{f}_j(T_n)$. Consequently, with same notations as in the previous paragraph, a subtree of T_n born from the root that appears with multiplicity 1 will also appear among $T_{(1)}, \ldots, T_{(p)}$.

Moreover, subtrees that are replaced are always replaced by trees with the same number of vertices and a tree with j vertices and edge-lengths a has height at most aj. So the Gromov-Hausdorff-Prokhorov distance between two trees with edge-lengths a that both decompose above the root in subtrees of same size j is at most $2aj$ (it is implicit in this proof that all trees are endowed with the uniform measure on their vertices). We now appeal to the following elementary
Fact. Let t, t' be such that $k = p(\lambda(\mathsf{t})) = p(\lambda(\mathsf{t}'))$, and let $\mathsf{t} = \langle \mathsf{t}_{(1)}, \ldots, \mathsf{t}_{(k)}\rangle$ and $\mathsf{t}' = \langle \mathsf{t}'_{(1)}, \ldots, \mathsf{t}'_{(k)}\rangle$ with $\#\mathsf{t}_{(i)} = \#\mathsf{t}'_{(i)}$ for $1 \leq i \leq k$. Then for every $a > 0$,

$$d_{\mathrm{GHP}}(a\mathsf{t}, a\mathsf{t}') \leq \max_{1 \leq i \leq k} d_{\mathrm{GHP}}(a\mathsf{t}_{(i)}, a\mathsf{t}'_{(i)}) .$$

From this we deduce that the Gromov-Hausdorff-Prokhorov distance between aT_n and aT'_n is at most

$$\left(2a \sup\{1 \leq j \leq n-1 : F_j \in A_j^c\}\right) \vee \sup_{\substack{1 \leq j \leq n-1 \\ F_j \in A_j}} \sup_{i : \#T_{(i)} = j} d_{\mathrm{GHP}}(aT_{(i)}, aT'_{(i)}) ,$$

where $(T_{(i)}, T'_{(i)})$ is the natural coupling. The induction hypothesis allows to conclude.

Last, for (iii), fix $\varepsilon \in (0, 1)$. The Gromov-Hausdorff-Prokhorov distance between $n^{-\varepsilon}T_n$ and $n^{-\varepsilon}T'_n$ is bounded from above by $2n^{1-\varepsilon}$ for all $n \geq 1$. Next, for $\gamma \in (0, 1)$, let A_n^γ be the subset of trees of $\mathsf{T}_n^{(m)}$ that have at least two subtrees born from the same vertex that are identical, and with size larger than n^γ. By (ii), when $T_n \notin A_n^\gamma$, $d_{\mathrm{GHP}}(n^{-\varepsilon}T_n, n^{-\varepsilon}T'_n) \leq 2n^{\gamma-\varepsilon}$. Hence,

$$\begin{aligned} \mathbb{E}\left[d_{\mathrm{GHP}}(n^{-\varepsilon}T_n, n^{-\varepsilon}T'_n)\right] &= \mathbb{E}\left[d_{\mathrm{GHP}}(n^{-\varepsilon}T_n, n^{-\varepsilon}T'_n)\mathbf{1}_{\{T_n \in A_n^\gamma\}}\right] + \mathbb{E}\left[d_{\mathrm{GHP}}(n^{-\varepsilon}T_n, n^{-\varepsilon}T'_n)\mathbf{1}_{\{T_n \notin A_n^\gamma\}}\right] \\ &\leq 2n^{1-\varepsilon}\mathbb{P}(T_n \in A_n^\gamma) + 2n^{\gamma-\varepsilon}. \end{aligned}$$

Taking $\gamma < \varepsilon$ and using Lemma 24 following right below, we get the result. $\square$

Lemma 24. *For $\gamma \in (0,1)$, let A_n^γ be the subset of $\mathsf{T}_n^{(m)}$ of trees t that have at least one vertex v such that at least two subtrees born from v are equal and have at least n^γ vertices. Then,*

$$\mathbb{P}(T_n \in A_n^\gamma) = O(\rho^{-n^\gamma} n^{5/2}) \text{ as } n \to \infty.$$

This lemma will be an easy consequence of the following result. For every tree t and any vertex v of t, we let $\mathsf{t}^{(v)}$ denote the subtree of t rooted at v. When v^* is taken uniformly at random among the vertices of t, we set $\mathsf{t}^{(*)} := \mathsf{t}^{(v^*)}$.

Lemma 25. *The distribution of $T_n^{(*)}$ conditionally on $\#T_n^{(*)} = k$ is uniform on $\mathsf{T}_k^{(m)}$, for every $1 \leq k \leq n$.*

Note that the event $\{\#T_n^{(*)} = k\}$ has a strictly positive probability for all $1 \leq k \leq n$.

Proof. Let $k \in \{1, \ldots, n\}$. For all $\mathsf{t}_0 \in \mathsf{T}_k^{(m)}$, using that $\mathbb{P}(T_n = \mathsf{t}) = 1/\mathbf{T}_n^{(m)}$ for $\mathsf{t} \in \mathsf{T}_n^{(m)}$,

$$\begin{aligned}
\mathbb{P}\big(T_n^{(*)} = \mathsf{t}_0\big) &= \sum_{\mathsf{t} \in \mathsf{T}_n^{(m)}} \mathbb{P}\big(T_n^{(*)} = \mathsf{t}_0 | T_n = \mathsf{t}\big) \mathbb{P}(T_n = \mathsf{t}) \\
&= \frac{1}{\mathbf{T}_n^{(m)}} \sum_{\mathsf{t} \in \mathsf{T}_n^{(m)}} \mathbb{P}(\mathsf{t}^{(*)} = \mathsf{t}_0) \\
&= \frac{1}{n\mathbf{T}_n^{(m)}} \sum_{\mathsf{t} \in \mathsf{T}_n^{(m)}, v \in \mathsf{t}} \mathbf{1}_{\{\mathsf{t}^{(v)} = \mathsf{t}_0\}}.
\end{aligned}$$

This quantity is independent of $\mathsf{t}_0 \in \mathsf{T}_k^{(m)}$, because there is an obvious bijection between the sets $\{(\mathsf{t}, v) : \mathsf{t} \in \mathsf{T}_n^{(m)}, v \in \mathsf{t}, \mathsf{t}^{(v)} = \mathsf{t}_0\}$ and $\{(\mathsf{t}, v) : \mathsf{t} \in \mathsf{T}_n^{(m)}, v \in \mathsf{t}, \mathsf{t}^{(v)} = \mathsf{t}_1\}$ for $\mathsf{t}_1 \in \mathsf{T}_k^{(m)}$. Hence the result. $\square$

Proof of Lemma 24. Let $A_n^\gamma(k)$ be the subset of trees of $\mathsf{T}_k^{(m)}$ whose decomposition above the root gives birth to at least two identical subtrees with size greater than n^γ, $k \leq n$. We first give an upper bound for the probability $\mathbb{P}(T_k \in A_n^\gamma(k))$. To do so, we bound from above the number of trees of $\mathsf{T}_k^{(m)}$ that decompose in at least two identical subtrees of size i ($i \leq (k-1)/2$): there are $\mathbf{T}_i^{(m)}$ choices for the tree with size i appearing twice. Then, there are $\mathbf{T}_{k-2i}^{(m)}$ forests with $k-1-2i$ vertices. Gluing the twins trees and a forest with $k-1-2i$ vertices to a common root gives a tree with k vertices (and a root branching in possibly more than m subtrees) and all trees in $\mathsf{T}_k^{(m)}$ with at least two subtrees with size i can be obtained like this. From this we deduce that the cardinal of $A_n^\gamma(k)$ is at most $\sum_{i=n^\gamma}^{(k-1)/2} \mathbf{T}_i^{(m)} \mathbf{T}_{k-2i}^{(m)}$. In particular, using (35) and the fact that $\rho > 1$,

$$\mathbb{P}\left(T_k \in A_n^\gamma(k)\right) \leq \frac{1}{\mathbf{T}_k^{(m)}} \sum_{i=n^\gamma}^{(k-1)/2} \mathbf{T}_i^{(m)} \mathbf{T}_{k-2i}^{(m)} \leq C \frac{k^{3/2}}{\rho^k} \sum_{i=n^\gamma}^{(k-1)/2} \frac{\rho^i \rho^{k-2i}}{i^{3/2}} \leq C n^{3/2} \rho^{-n^\gamma},$$

where C is a generic constant independent of n and $k \leq n$. Now, in the following lines, given T_n, we let $v_1, v_2, \ldots, v_n$ denote its vertices labeled uniformly at random,

$$\begin{aligned}
\mathbb{P}\left(T_n \in A_n^\gamma\right) &\leq \mathbb{E}\left[\sum_{v \in T_n} \mathbf{1}_{\{T_n^{(v)} \in A_n^\gamma(\#T_n^{(v)})\}}\right] = \mathbb{E}\left[\sum_{i=1}^n \mathbf{1}_{\{T_n^{(v_i)} \in A_n^\gamma(\#T_n^{(v_i)})\}}\right] \\
&= n\mathbb{P}\left(T_n^{(*)} \in A_n^\gamma(\#T_n^{(*)})\right) \\
&\underset{\text{by Lemma 25}}{=} n \sum_{k=1}^n \mathbb{P}\left(T_k \in A_n^\gamma(k)\right) \mathbb{P}\big(\#T_n^{(*)} = k\big) \\
&\leq C n^{5/2} \rho^{-n^\gamma} \sum_{k=1}^n \mathbb{P}\big(\#T_n^{(*)} = k\big) = C n^{5/2} \rho^{-n^\gamma}.
\end{aligned}$$

$\square$

6.2 Hypothesis (H) and conclusion

It remains to check that the family of probability distributions on $\mathcal{P}_n, n \geq 1$ defined by

$$q_n(\lambda) = \mathbb{P}\left(\lambda(T_{n+1}) = \lambda\right) = \mathbb{P}\left(\lambda(T'_{n+1}) = \lambda\right) = \frac{\mathbf{S}_{n+1}^{(\lambda)}}{\mathbf{T}_{n+1}^{(m)}}, \quad \forall \lambda \in \mathcal{P}_n,$$

satisfies the assumption **(H)** with $\gamma = 1/2$, $\ell \equiv 1$ and ν proportional to the Brownian dislocation measure ν_2. For this we recall and fix some more notations:

* $\widetilde{\mathsf{T}}_n^{(m)}$ is the subset of $\mathsf{T}_n^{(m)}$ of trees with root degree less or equal to $m-2$
* $\widetilde{\mathbf{T}}_n^{(m)}$ is the cardinal of $\widetilde{\mathsf{T}}_n^{(m)}$, and $\psi^{(m)}(x) = \sum_{n\geq 1} \mathbf{T}_n^{(m)} x^n$, $\widetilde{\psi}^{(m)}(x) = \sum_{n \geq 1} \widetilde{\mathbf{T}}_n^{(m)} x^n$
* for $\lambda = (\lambda_1, \lambda_2, \dots) \in \mathcal{P}_n$, set $\lambda_{\mathrm{r}} := \sum_{i=3}^{\infty} \lambda_i = n - \lambda_1 - \lambda_2$

The main result of this section then reads

Proposition 18. *For all $2 \leq m \leq \infty$, and all continuous functions $f : \mathcal{S}^{\downarrow} \to \mathbb{R}$ such that $|f(\mathbf{s})| \leq 1 - s_1$ for $\mathbf{s} \in \mathcal{S}^{\downarrow}$,*

$$\sqrt{n} \sum_{\lambda \in \mathcal{P}_n} f\left(\frac{\lambda}{n}\right) \frac{\mathbf{S}_{n+1}^{(\lambda)}}{\mathbf{T}_{n+1}^{(m)}} \underset{n\to\infty}{\longrightarrow} \kappa \widetilde{\psi}^{(m)}(1/\rho) \int_{1/2}^{1} \frac{f(x, 1-x, 0, \dots)}{x^{3/2}(1-x)^{3/2}} \mathrm{d}x.$$

Note that the sum in the limit above is finite, since $\widetilde{\mathbf{T}}_k^{(m)} \leq \mathbf{T}_k^{(m)} \leq K\rho^k/k^{3/2}$. This sum is explicit in terms of κ and ρ when $m=2$ or $m=\infty$. See Section 2.2 for details.

With this proposition, it is easy to conclude the proof of Theorem 4. Indeed, together with Theorem 2 and Proposition 17 (i), it leads to the convergence

$$\frac{1}{\sqrt{n}} T'_n \underset{n\to\infty}{\overset{(\mathrm{d})}{\longrightarrow}} c_m \mathcal{T}_{1/2,\nu_2}$$

for the Gromov-Hausdorff-Prokhorov topology, where $c_m = \sqrt{2}/(\sqrt{\pi}\kappa\widetilde{\psi}(1/\rho))$. Then, by Proposition 17 (iii) and since $(\mathcal{M}_{\mathrm{w}}, d_{\mathrm{GH}})$ is a complete separable space, we can apply a Slutsky-type theorem to get

$$\frac{1}{\sqrt{n}} T_n \underset{n\to\infty}{\overset{(\mathrm{d})}{\longrightarrow}} c_m \mathcal{T}_{1/2,\nu_2}.$$

The rest of this section is devoted to the proof of Proposition 18.

6.2.1 Negligible terms

We show in this section that the set of partitions $\lambda \in \mathcal{P}_n$ where either $\lambda_1 \geq n(1-\varepsilon)$ or $\lambda_{\mathrm{r}} \geq n\varepsilon$ plays a negligible role in the limit of Proposition 18 when we first let $n \to \infty$ and then $\varepsilon \to 0$.

Lemma 26. *There exists $C \in (0,\infty)$ such that, for all $0 < \varepsilon < 1$,*

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{\mathbf{T}_{n+1}^{(m)}} \sum_{\lambda \in \mathcal{P}_n} \mathbf{1}_{\{\lambda_1 \geq n(1-\varepsilon)\}} \left(1 - \frac{\lambda_1}{n}\right) \mathbf{S}_{n+1}^{(\lambda)} \leq \frac{C\sqrt{\varepsilon}}{(1-\varepsilon)^{3/2}}.$$

Proof. Using (34) and then (35), we get

$$\begin{aligned}\sum_{\lambda\in\mathcal{P}_n,\lambda_1\geq n(1-\varepsilon)}\left(1-\frac{\lambda_1}{n}\right)\mathbf{S}_{n+1}^{(\lambda)} &\leq \sum_{\lambda_1=\lceil n(1-\varepsilon)\rceil}^{n}\left(1-\frac{\lambda_1}{n}\right)\mathbf{T}_{\lambda_1}^{(m)}\sum_{\mu\in\mathcal{P}_{n-\lambda_1}}\mathbf{S}_{n+1-\lambda_1}^{(\mu)}\\ &\leq \sum_{\lambda_1=\lceil n(1-\varepsilon)\rceil}^{n}\left(1-\frac{\lambda_1}{n}\right)\mathbf{T}_{\lambda_1}^{(m)}\mathbf{T}_{n+1-\lambda_1}^{(m)}\\ &\leq K^2\rho^{n+1}\sum_{\lambda_1=\lceil n(1-\varepsilon)\rceil}^{n-1}\frac{1-\frac{\lambda_1}{n}}{\lambda_1^{3/2}(n+1-\lambda_1)^{3/2}}\\ &\leq \frac{K^2\rho^{n+1}}{(n(1-\varepsilon))^{3/2}}\times\frac{1}{n^{3/2}}\times\sum_{\lambda_1=\lceil n(1-\varepsilon)\rceil}^{n-1}\frac{1}{(1-\lambda_1/n)^{1/2}}.\end{aligned}$$

We conclude with the fact that the sum $\sum_{\lambda_1=\lceil n(1-\varepsilon)\rceil}^{n-1}(1-\lambda_1/n)^{-1/2}$ is smaller than the integral $\int_{n(1-\varepsilon)}^{n}(1-x/n)^{-1/2}\mathrm{d}x = 2n\sqrt{\varepsilon}$, and then use the lower bound of (35) for $\mathbf{T}_{n+1}^{(m)}$. □

To deal with the partitions where $\lambda_{\mathrm{r}} \geq n\varepsilon$, we need the following lemma when $m=\infty$. We denote by $\mathbf{T}_n^{(\infty,a-)}$ the number of trees of $\mathsf{T}_n^{(\infty)}$ whose subtrees born from the root have sizes at most a, $a\geq 1$.

Lemma 27. *Let $m=\infty$. There exists $A,B>0$ such that*

$$\mathbf{T}_{k+1}^{(\infty,a-)} \leq A\rho^k\exp(-Bk/a), \quad \forall k\in\mathbb{N} \text{ and } a\geq 1.$$

Proof. Recall that T denotes the set of all (rooted, unordered) trees and rewrite the power series $\psi=\psi^{(\infty)}$ as $\psi(x)=\sum_{\mathsf{t}\in\mathsf{T}}x^{\#\mathsf{t}}$. According to [23, Section VII.5] its radius of convergence is $1/\rho<1$ and $\psi(1/\rho)=1$. Note also that $\psi(0)=0$. Now, we consider a random tree T in T with distribution defined by

$$\mathbb{P}(T=\mathsf{t})=\rho^{-\#\mathsf{t}}.$$

If $c_\varnothing(\mathsf{t})$ denotes the degree of the root of t, we just have to show that

$$\mathbb{P}(c_\varnothing(T)=r)\leq A'\exp(-B'r), \text{ for some } A',B'>0 \text{ and all } r\geq 1. \tag{36}$$

Indeed, each tree with $k+1$ vertices and a decomposition in subtrees with sizes at most a has a root degree larger or equal to k/a. So, if the above inequality holds, we will have

$$\mathbf{T}_{k+1}^{(\infty,a-)}\leq \rho^{k+1}\mathbb{P}\big(c_\varnothing(T)\geq k/a, T\in\mathsf{T}_{k+1}^{(\infty)}\big)\leq \rho^{k+1}A'B'^{-1}\exp(-B'(k/a-1)),$$

as required. To get (36), note that

$$\mathbb{P}(c_\varnothing(T)=r)=\sum_{\mathsf{t}\in\mathsf{T},c_\varnothing(\mathsf{t})=r}\rho^{-\#\mathsf{t}}=\sum_{k=1}^{r}\frac{1}{k!}\sum_{\substack{\mathsf{t}_1,\ldots,\mathsf{t}_k\in\mathsf{T}\\ \text{pairwise distinct}}}\sum_{\substack{m_1+\ldots+m_k=r\\ m_i\geq 1}}\rho^{-1-\sum_{1\leq i\leq k}m_i\#t_i}$$

which is obtained by considering the multiset of r subtrees of a tree t, made of k distinct trees with multiplicities $m_1,\ldots,m_k$. Hence,

$$\begin{aligned}\mathbb{P}(c_\varnothing(T)=r) &\leq \rho^{-1}\sum_{k=1}^{r}\frac{1}{k!}\sum_{\substack{m_1+\ldots+m_k=r\\ m_i\geq 1}}\prod_{i=1}^{k}\psi(\rho^{-m_i})\\ &= \rho^{-1}\sum_{k=1}^{\lfloor cr\rfloor}\frac{1}{k!}\sum_{\substack{m_1+\ldots+m_k=r\\ m_i\geq 1}}\prod_{i=1}^{k}\psi(\rho^{-m_i})+\rho^{-1}\sum_{k=\lfloor cr\rfloor+1}^{r}\frac{1}{k!}\sum_{\substack{m_1+\ldots+m_k=r\\ m_i\geq 1}}\prod_{i=1}^{k}\psi(\rho^{-m_i})\end{aligned}$$

where the $c \in]0,1[$ chosen for this split will be specified below.

We first bound from above the second term. Using that $\psi(\rho^{-m_i}) \leq \psi(\rho^{-1}) = 1$ for $m_i \geq 1$ and that $\sum_{m_1+\ldots+m_k=r, m_i\geq 1} = \binom{r-1}{k-1}$, we obtain

$$\sum_{k=\lfloor cr\rfloor+1}^{r} \frac{1}{k!} \sum_{\substack{m_1+\ldots+m_k=r\\ m_i\geq 1}} \prod_{i=1}^{k} \psi(\rho^{-m_i}) \leq \frac{1}{\lfloor cr\rfloor!} \sum_{k=1}^{r} \binom{r-1}{k-1} \leq \frac{2^{r-1}}{\lfloor cr\rfloor!},$$

which decays exponentially fast as $r \to \infty$, for every $c \in]0,1[$.

Now we will check that the sum $\sum_{k=1}^{\lfloor cr\rfloor} \frac{1}{k!} \sum_{m_1+\ldots+m_k=r, m_i\geq 1} \prod_{i=1}^{k} \psi(\rho^{-m_i})$ also decays exponentially in r, provided that $c \in]0,1[$ is chosen sufficiently small. Since $\psi(0) = 0$, we have that $\psi(x) \leq Cx$ for some $C < \infty$ and all $x \in [0, \rho^{-1}]$. Hence

$$\begin{aligned}
\sum_{k=1}^{\lfloor cr\rfloor} \frac{1}{k!} \sum_{\substack{m_1+\ldots+m_k=r\\ m_i\geq 1}} \prod_{i=1}^{k} \psi(\rho^{-m_i}) \quad & \leq \quad \sum_{k=1}^{\lfloor cr\rfloor} \frac{C^k}{k!} \sum_{\substack{m_1+\ldots+m_k=r\\ m_i\geq 1}} \prod_{i=1}^{k} \rho^{-m_i} \\
& \leq \quad \exp(C) \sum_{k=1}^{\lfloor cr\rfloor} \rho^{-r} \binom{r-1}{k-1} \\
& \underset{\text{for all } \lambda>0}{\leq} \quad \exp(C)\rho^{-r} \sum_{k=r-\lfloor cr\rfloor}^{r-1} \binom{r-1}{k} \exp\left(\lambda k - \lambda(r - \lfloor cr\rfloor)\right) \\
& \leq \quad \exp(C) \left(\rho^{-1}\exp(-\lambda(1-c))(\exp(\lambda)+1)\right)^r .
\end{aligned}$$

When $c \to 0$, $\rho^{-1}\exp(-\lambda(1-c))(\exp(\lambda)+1) \to \rho^{-1}(1+\exp(-\lambda))$, which is strictly smaller than 1 for λ large enough. Hence, fix such a large λ and then take $c > 0$ sufficiently small so that $\rho^{-1}\exp(-\lambda(1-c))(\exp(\lambda)+1) < 1$. This ends the proof. $\square$

Lemma 28. *For all $\varepsilon > 0$,*

$$\frac{\sqrt{n}}{\mathbf{T}_{n+1}^{(m)}} \sum_{\lambda\in\mathcal{P}_n} \mathbf{1}_{\{\lambda_{\mathrm{r}}\geq n\varepsilon\}} \mathbf{S}_{n+1}^{(\lambda)} \underset{n\to\infty}{\longrightarrow} 0.$$

Proof. • If $m = 2$, $\lambda_{\mathrm{r}} = 0$ for all $\lambda \in \mathcal{P}_n$ and the assertion is obvious.

• Assume now that $3 \leq m < \infty$ and note that when $\lambda \in \mathcal{P}_n$ with $p(\lambda) \leq m$, one has that $\lambda_{\mathrm{r}} \geq n\varepsilon$ implies $(m-2)\lambda_3 \geq n\varepsilon$, in particular $\lambda_1 \geq \lambda_2 \geq n\varepsilon/(m-2)$. Hence,

$$\sum_{\lambda\in\mathcal{P}_n} \mathbf{1}_{\{\lambda_{\mathrm{r}}\geq n\varepsilon\}} \mathbf{S}_{n+1}^{(\lambda)} \leq \sum_{\lambda_{\mathrm{r}}=\lceil n\varepsilon\rceil}^{n-2} \sum_{\lambda_1=\lceil n\varepsilon/(m-2)\rceil}^{\lfloor n-\lambda_{\mathrm{r}}-n\varepsilon/(m-2)\rfloor^+} \mathbf{T}_{\lambda_1}^{(m)} \mathbf{T}_{n-\lambda_{\mathrm{r}}-\lambda_1}^{(m)} \mathbf{T}_{\lambda_{\mathrm{r}}+1}^{(m)}.$$

Then for C a generic constant, using (35), the latter term multiplied by $\sqrt{n}/\mathbf{T}_{n+1}^{(m)}$ is bounded from above by

$$\begin{aligned}
& Cn^{1/2}(n+1)^{3/2} \sum_{\lambda_{\mathrm{r}}=\lceil n\varepsilon\rceil}^{n-2} \sum_{\lambda_1=\lceil n\varepsilon/(m-2)\rceil}^{\lfloor n-\lambda_{\mathrm{r}}-n\varepsilon/(m-2)\rfloor^+} \frac{1}{\lambda_1^{3/2}(n-\lambda_{\mathrm{r}}-\lambda_1)^{3/2}(\lambda_{\mathrm{r}}+1)^{3/2}} \\
\leq \quad & C\frac{n^2(n-2)(\lfloor n-n\varepsilon/(m-2)\rfloor^+)}{n^{3*3/2}} = O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

• Next we turn to the case where $m = \infty$. Let $\gamma \in (5/6, 1)$. On the one hand, by the same token as for the $m < \infty$ cases,

$$\begin{aligned}\frac{\sqrt{n}}{\mathbf{T}^{(\infty)}_{n+1}} \sum_{\lambda \in \mathcal{P}_n} \mathbf{1}_{\{\lambda_{\mathrm{r}} \geq n\varepsilon\}} \mathbf{1}_{\{\lambda_2 \geq n^\gamma\}} \mathbf{S}^{(\lambda)}_{n+1} &\leq Cn^2 \sum_{\lambda_{\mathrm{r}} = \lceil n\varepsilon \rceil}^{n-2} \sum_{\lambda_1 = \lceil n^\gamma \rceil}^{\lfloor n - \lambda_{\mathrm{r}} - n^\gamma \rfloor^+} \frac{1}{\lambda_1^{3/2} (n - \lambda_{\mathrm{r}} - \lambda_1)^{3/2} (\lambda_{\mathrm{r}} + 1)^{3/2}} \\ &\leq C \frac{n^4}{n^{3\gamma + 3/2}} = O\left(n^{5/2 - 3\gamma}\right) = o(1),\end{aligned}$$

since $5/2 - 3\gamma < 0$ when $\gamma > 5/6$. On the other hand, we use Lemma 27 to get

$$\begin{aligned}\frac{\sqrt{n}}{\mathbf{T}^{(\infty)}_{n+1}} \sum_{\lambda \in \mathcal{P}_n} \mathbf{1}_{\{\lambda_{\mathrm{r}} \geq n\varepsilon\}} \mathbf{1}_{\{\lambda_2 < n^\gamma\}} \mathbf{S}^{(\lambda)}_{n+1} &\leq \frac{\sqrt{n}}{\mathbf{T}^{(\infty)}_{n+1}} \sum_{\lambda_{\mathrm{r}} = \lceil n\varepsilon \rceil}^{n-2} \sum_{\lambda_1 = 1}^{n - \lambda_{\mathrm{r}} - 1} \mathbf{T}^{(\infty)}_{\lambda_1} \mathbf{T}^{(\infty)}_{n - \lambda_{\mathrm{r}} - \lambda_1} \mathbf{T}^{(\infty, n^\gamma -)}_{\lambda_{\mathrm{r}} + 1} \\ &\leq Cn^4 \exp(-Bn^{1-\gamma}\varepsilon) = o(1).\end{aligned}$$

□

6.2.2 Proof of Proposition 18

We rely on the following lemma. Let $\mathcal{P}^{\mathrm{bin}}_n$ be the subset of $\mathcal{P}_n$ of partitions of n with exactly two parts.

Lemma 29. *Let $f : \mathcal{S}^{\downarrow} \to \mathbb{R}$ be continuous.*
(i) *For all $a \in \mathbb{Z}_+$ and all $\varepsilon \in (0,1)$, as $n \to \infty$,*

$$\frac{\sqrt{n}}{\mathbf{T}^{(m)}_{n+1}} \sum_{\substack{\lambda \in \mathcal{P}^{\mathrm{bin}}_{n-a} \\ \lambda_1 \leq n(1-\varepsilon)}} f\left(\frac{\lambda_1}{n}, \frac{\lambda_2 + a}{n}, 0, \ldots\right) \mathbf{S}^{(\lambda)}_{n+1-a} \longrightarrow \frac{\kappa}{\rho^{1+a}} \int_{1/2}^{1-\varepsilon} \frac{f(x, 1-x, 0, \ldots)}{x^{3/2}(1-x)^{3/2}} \mathrm{d}x.$$

(ii) *Moreover, there exists $C_\varepsilon \in (0,\infty)$ such that, for all $n \geq 1$, all $0 \leq a \leq n\varepsilon/2$ and all non-increasing non-negative sequences $(a_i, i \geq 1)$ with $\sum_{i \geq 1} a_i \leq a/n$,*

$$\left| \frac{\sqrt{n}}{\mathbf{T}^{(m)}_{n+1}} \sum_{\substack{\lambda \in \mathcal{P}^{\mathrm{bin}}_{n-a} \\ \lambda_1 \leq n(1-\varepsilon)}} f\left(\frac{\lambda_1}{n}, \frac{\lambda_2}{n} + a_1, a_2, a_3, \ldots\right) \mathbf{S}^{(\lambda)}_{n+1-a} \mathbf{T}^{(m)}_{a+1} \right| \leq \frac{C_\varepsilon}{(a+1)^{3/2}}.$$

Proof. (i) For large enough n,

$$\begin{aligned}\sum_{\substack{\lambda \in \mathcal{P}^{\mathrm{bin}}_{n-a} \\ \lambda_1 \leq n(1-\varepsilon)}} f\left(\frac{\lambda_1}{n}, \frac{\lambda_2 + a}{n}, 0, \ldots\right) \mathbf{S}^{(\lambda)}_{n+1-a} &= f\left(\frac{1}{2} - \frac{a}{2n}, \frac{1}{2} + \frac{a}{2n}, 0, \ldots\right) \mathbf{F}_{(n-a)/2}(2) \mathbf{1}_{\{n-a \text{ is even}\}} \\ &+ \sum_{\lambda_1 = \lfloor (n-a)/2 \rfloor + 1}^{\lfloor n(1-\varepsilon) \rfloor} f\left(\frac{\lambda_1}{n}, 1 - \frac{\lambda_1}{n}, 0, \ldots\right) \mathbf{T}^{(m)}_{\lambda_1} \mathbf{T}^{(m)}_{n-a-\lambda_1}.\end{aligned}$$

On the one hand, by Otter's approximation result for $\mathbf{T}^{(m)}_{(n-a)/2}$

$$\mathbf{F}_{(n-a)/2}(2) = \mathbf{T}^{(m)}_{(n-a)/2} \big(\mathbf{T}^{(m)}_{(n-a)/2} + 1\big)/2 \sim \kappa^2 \rho^{n-a} ((n-a)/2)^{-3}/2 = o(\mathbf{T}^{(m)}_{n+1}/\sqrt{n}).$$

On the other hand, still using Otter's result, we get that for all $\eta > 0$, provided that n is large enough,

$$
\begin{aligned}
& \frac{\sqrt{n}}{\mathbf{T}^{(m)}_{n+1}} \sum_{\lambda_1=\lfloor (n-a)/2\rfloor+1}^{\lfloor n(1-\varepsilon)\rfloor} f\left(\frac{\lambda_1}{n}, 1-\frac{\lambda_1}{n}, 0, \ldots\right) \mathbf{T}^{(m)}_{\lambda_1}\mathbf{T}^{(m)}_{n-a-\lambda_1} \\
\leq \quad & \frac{(\kappa+\eta)^2}{(\kappa-\eta)\rho^{1+a}} \frac{1}{n} \sum_{\lambda_1=\lfloor (n-a)/2\rfloor+1}^{\lfloor n(1-\varepsilon)\rfloor} f\left(\frac{\lambda_1}{n}, 1-\frac{\lambda_1}{n}, 0, \ldots\right) \frac{(n+1)^{3/2}}{\lambda_1^{3/2}} \times \frac{n^{3/2}}{(n-a-\lambda_1)^{3/2}} \\
\underset{n\to\infty}{\longrightarrow} \quad & \frac{(\kappa+\eta)^2}{(\kappa-\eta)\rho^{1+a}} \int_{1/2}^{1-\varepsilon} \frac{f(x,1-x,0,\ldots)}{x^{3/2}(1-x)^{3/2}} \mathrm{d}x.
\end{aligned}
$$

Letting $\eta \to 0$, this gives

$$
\limsup_{n\to\infty} \frac{\sqrt{n}}{\mathbf{T}^{(m)}_{n+1}} \sum_{\lambda_1=\lfloor (n-a)/2\rfloor+1}^{\lfloor n(1-\varepsilon)\rfloor} f\left(\frac{\lambda_1}{n}, 1-\frac{\lambda_1}{n}, 0, \ldots\right) \mathbf{T}^{(m)}_{\lambda_1}\mathbf{T}^{(m)}_{n-a-\lambda_1} \leq \frac{\kappa}{\rho^{1+a}} \int_{1/2}^{1-\varepsilon} \frac{f(x,1-x,0,\ldots)}{x^{3/2}(1-x)^{3/2}} \mathrm{d}x.
$$

We obtain the liminf similarly, hence (i).

(ii) We will use that $\mathbf{S}^{(\lambda)}_{n+1-a} \leq \mathbf{T}^{(m)}_{\lambda_1}\mathbf{T}^{(m)}_{n-a-\lambda_1}$ for all $\lambda \in \mathcal{P}^{\mathrm{bin}}_{n-a}$. Recall that f is bounded on $\mathcal{S}^{\downarrow}$. There exits then a generic constant C independent of n and $a \leq n\varepsilon/2$ such that

$$
\begin{aligned}
& \left| \frac{\sqrt{n}}{\mathbf{T}^{(m)}_{n+1}} \sum_{\lambda\in\mathcal{P}^{\mathrm{bin}}_{n-a}, \lambda_1\leq n(1-\varepsilon)} f\left(\frac{\lambda_1}{n}, \frac{\lambda_2}{n}+a_1, a_2, a_3, \ldots\right) \mathbf{S}^{(\lambda)}_{n+1-a}\mathbf{T}^{(m)}_{a+1} \right| \\
\leq \quad & \frac{C\sqrt{n}}{\mathbf{T}^{(m)}_{n+1}} \sum_{\lambda_1=\lceil (n-a)/2\rceil}^{\lfloor n(1-\varepsilon)\rfloor} \mathbf{T}^{(m)}_{\lambda_1}\mathbf{T}^{(m)}_{n-a-\lambda_1}\mathbf{T}^{(m)}_{a+1} \\
\leq \quad & \frac{C}{(a+1)^{3/2}} \frac{1}{n} \sum_{\lambda_1=\lceil (n-a)/2\rceil}^{\lfloor n(1-\varepsilon)\rfloor} \frac{(n+1)^{3/2}}{\lambda_1^{3/2}} \times \frac{n^{3/2}}{(n-a-\lambda_1)^{3/2}} \\
\leq \quad & \frac{C}{(a+1)^{3/2}} \frac{1}{n} \sum_{\lambda_1=\lceil n(1-\varepsilon/2)/2\rceil}^{\lfloor n(1-\varepsilon)\rfloor} \frac{n^{3/2}}{\lambda_1^{3/2}} \times \frac{n^{3/2}}{(n-\lambda_1)^{3/2}}
\end{aligned}
$$

where we have used for the last inequality that $n-a \geq n(1-\varepsilon/2)$ since $a \leq n\varepsilon/2$ and that $n-a-\lambda_1 \geq (n-\lambda_1)/2$ since $a \leq n\varepsilon/2$ and $\lambda_1 \leq n(1-\varepsilon)$. This upper bound is of the form $Cu_n/(a+1)^{3/2}$ where $(u_n, n \geq 1)$ is a sequence independent of a converging to a finite limit as $n \to \infty$. Hence the result. $\square$

Proof of Proposition 18. By Lemmas 26 and 28, the set of partitions where either $\lambda_1 \geq n(1-\varepsilon)$ or $\lambda_{\mathrm{r}} \geq n\varepsilon/3$ will play a negligible role in the limit when we first let $n \to \infty$ and then $\varepsilon \to 0$. Hence we concentrate first on the following sums (for $\varepsilon \in (0,1)$), where we use that for all $\lambda \in \mathcal{P}_n$, $\lambda_1 \leq n(1-\varepsilon)$ and $\lambda_{\mathrm{r}} \leq n\varepsilon/3$ implies $\lambda_2 > \lambda_3$:

$$
\begin{aligned}
\sum_{\substack{\lambda\in\mathcal{P}_n \\ \lambda_1\leq n(1-\varepsilon), \lambda_{\mathrm{r}}\leq n\varepsilon/3}} f\left(\frac{\lambda}{n}\right) \mathbf{S}^{(\lambda)}_{n+1} &= \sum_{k=0}^{\lfloor n\varepsilon/3\rfloor} \sum_{\substack{\mu\in\mathcal{P}_k \\ p(\mu)\leq m-2}} \sum_{\substack{\lambda\in\mathcal{P}^{\mathrm{bin}}_{n-k} \\ \lambda_1\leq n(1-\varepsilon)}} f\left(\frac{\lambda_1}{n}, \frac{\lambda_2+k}{n}, 0, \ldots\right) \mathbf{S}^{(\lambda)}_{n-k+1}\mathbf{S}^{(\mu)}_{k+1} \qquad (37) \\
&+ \sum_{k=0}^{\lfloor n\varepsilon/3\rfloor} \sum_{\substack{\mu\in\mathcal{P}_k \\ p(\mu)\leq m-2}} \sum_{\substack{\lambda\in\mathcal{P}^{\mathrm{bin}}_{n-k} \\ \lambda_1\leq n(1-\varepsilon)}} \left(f\left(\frac{\lambda_1}{n}, \frac{\lambda_2}{n}, \frac{\mu_1}{n}, \ldots\right) - f\left(\frac{\lambda_1}{n}, \frac{\lambda_2+k}{n}, 0, \ldots\right)\right) \mathbf{S}^{(\lambda)}_{n-k+1}\mathbf{S}^{(\mu)}_{k+1}.
\end{aligned}
$$

The first sum in the right-hand side of (37) is equal to

$$\sum_{k=0}^{\lfloor n\varepsilon/3\rfloor} \sum_{\lambda\in\mathcal{P}_{n-k}^{\mathrm{bin}},\lambda_1\le n(1-\varepsilon)} f\left(\frac{\lambda_1}{n},\frac{\lambda_2+k}{n},0,\dots\right)\mathbf{S}_{n-k+1}^{(\lambda)}\tilde{\mathbf{T}}_{k+1}^{(m)}, \tag{38}$$

which, multiplied by $\sqrt{n}/\mathbf{T}_{n+1}^{(m)}$, according to Lemma 29 (i) and (ii) ((ii) implies dominated convergence), converges to

$$\sum_{k=0}^{\infty}\widetilde{\mathbf{T}}_{k+1}^{(m)}\frac{\kappa}{\rho^{1+k}}\int_{1/2}^{1-\varepsilon}\frac{f(x,1-x,0,\dots)}{x^{3/2}(1-x)^{3/2}}\mathrm{d}x. \tag{39}$$

Next, let $\delta>0$. Since f is continuous (hence uniformly continuous) on the compact set $\mathcal{S}^{\downarrow}$, we can choose ε small enough so that the absolute value of the second sum in the right-hand side of (37) is bounded from above by

$$2\sum_{k=0}^{\lfloor n\varepsilon/3\rfloor} \sum_{\lambda\in\mathcal{P}_{n-k}^{\mathrm{bin}},\lambda_1\le n(1-\varepsilon)}\left(\delta\wedge\left(1-\frac{\lambda_1}{n}\right)\right)\mathbf{S}_{n-k+1}^{(\lambda)}\widetilde{\mathbf{T}}_{k+1}^{(m)}. \tag{40}$$

Similarly as above, when multiplied by $\sqrt{n}/\mathbf{T}_{n+1}^{(m)}$, this quantity converges to

$$2\sum_{k=0}^{\infty}\widetilde{\mathbf{T}}_{k+1}^{(m)}\frac{\kappa}{\rho^{1+k}}\int_{1/2}^{1-\varepsilon}\frac{\delta\wedge(1-x)}{x^{3/2}(1-x)^{3/2}}\mathrm{d}x\le 2\sum_{k=0}^{\infty}\widetilde{\mathbf{T}}_{k+1}^{(m)}\frac{\kappa}{\rho^{1+k}}\int_{1/2}^{1}\frac{\delta\wedge(1-x)}{x^{3/2}(1-x)^{3/2}}\mathrm{d}x \tag{41}$$

by Lemma 29 (i) and (ii).

Now let $\eta>0$ be fixed. For δ and ε sufficiently small, the terms (41) and the limsup of Lemma 26 are smaller than η, and the term (39) is in a neighborhood of radius η of the intended limit

$$\kappa\sum_{k=0}^{\infty}\frac{\widetilde{\mathbf{T}}_{k+1}^{(m)}}{\rho^{k+1}}\int_{1/2}^{1}\frac{f(x,1-x,0,\dots)}{x^{3/2}(1-x)^{3/2}}\mathrm{d}x. \tag{42}$$

Next, such small δ and ε being fixed, letting $n\to\infty$, and using Lemma 28 and the convergences of (38) to (39) and of (40) to (41), we get that $\sqrt{n}\sum_{\lambda\in\mathcal{P}_n} f\left(\frac{\lambda}{n}\right)\mathbf{S}_{n+1}^{(\lambda)}/\mathbf{T}_{n+1}^{(m)}$ is indeed in a neighborhood of radius 7η of (42) for all n large enough. $\square$

References

[1] D. Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.

[2] D. Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991.

[3] D. Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[4] D. Aldous. Probability distributions on cladograms. In *Random discrete structures (Minneapolis, MN, 1993)*, volume 76 of *IMA Vol. Math. Appl.*, pages 1–18. Springer, New York, 1996.

[5] D. J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985.

[6] J. Berestycki. Ranked fragmentations. *ESAIM Probab. Statist.*, 6:157–175 (electronic), 2002. Available via `http://www.emath.fr/ps/`.

[7] J. Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.

[8] J. Bertoin. Self-similar fragmentations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(3):319–340, 2002.

[9] J. Bertoin. *Random fragmentation and coagulation processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

[10] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999.

[11] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

[12] N. Broutin, L. Devroye, E. McLeish, and M. de la Salle. The height of increasing trees. *Random Structures Algorithms*, 32(4):494–518, 2008.

[13] N. Broutin and P. Flajolet. The height of random binary unlabelled trees. In *Fifth Colloquium on Mathematics and Computer Science*, Discrete Math. Theor. Comput. Sci. Proc., AI, pages 121–134. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2008.

[14] B. Chen, D. Ford, and M. Winkel. A new family of markov branching trees: the alpha-gamma model. `arXiv:0807.0554`, 2008.

[15] L. Devroye. A note on the height of binary search trees. *J. Assoc. Comput. Mach.*, 33(3):489–498, 1986.

[16] M. Drmota and B. Gittenberger. The shape of unlabeled rooted random trees. 2010. arXiv:1003.1322.

[17] T. Duquesne. A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.*, 31(2):996–1027, 2003.

[18] T. Duquesne and J.-F. Le Gall. Random trees, Lévy processes and spatial branching processes. *Astérisque*, (281):vi+147, 2002.

[19] T. Duquesne and J.-F. Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields*, 131(4):553–603, 2005.

[20] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[21] S. N. Evans, J. Pitman, and A. Winter. Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Related Fields*, 134(1):81–126, 2006.

[22] S. N. Evans and A. Winter. Subtree prune and regraft: a reversible real tree-valued Markov process. *Ann. Probab.*, 34(3):918–961, 2006.

[23] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[24] B. Haas and G. Miermont. The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electron. J. Probab.*, 9:no. 4, 57–97 (electronic), 2004.

[25] B. Haas and G. Miermont. Self-similar scaling limits of non-increasing Markov chains. 2009. arXiv:0909.3764.

[26] B. Haas, G. Miermont, J. Pitman, and M. Winkel. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *Ann. Probab.*, 36(5):1790–1837, 2008.

[27] B. Haas, J. Pitman, and M. Winkel. Spinal partitions and invariance under re-rooting of continuum random trees. *Ann. Probab.*, 37(4):1381–1411, 2009.

[28] J.-F. Le Gall and Y. Le Jan. Branching processes in Lévy processes: the exploration process. *Ann. Probab.*, 26(1):213–252, 1998.

[29] J.-F. Marckert and G. Miermont. The CRT is the scaling limit of unordered binary trees. *Random Structures and Algorithms*, 2010. To appear. arXiv:0902.4570.

[30] G. Miermont. Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér. (4)*, 42(5):725–781, 2009.

[31] R. Otter. The number of trees. *Ann. of Math. (2)*, 49:583–599, 1948.

[32] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Saint-Flour Summer School on Probability Theory.

[33] J. Pitman and M. Winkel. Regenerative tree growth: binary self-similar continuum random trees and poisson-dirichlet compositions. *Ann. Probab.*, 2009. To appear.